Trends in Symptom Validity, Memory and Psychological Test Performance as Functions of Time and Malingering Rating

A Thesis

Submitted to the Faculty

of

Drexel University

by

Maryellen McClain

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

September 2003

Dedication

This dissertation is dedicated to my mother, Terri, who has provided the inspiration for this, and many other, achievements throughout my life.

Acknowledgments

First, I would like to thank Doug Chute, Ph.D., my dissertation chair, graduate advisor, mentor and friend, for the support, direction, encouragement, and wisdom that he has provided me not only throughout this project, but also for the many years preceding it. I am especially grateful for his insights into the Zen of the Dissertation, and I can only hope that I have attained Dissertation Nirvana. I am also grateful to F. William Black, Ph.D., for his three years of mentorship, support, guidance and regular reports from Bill's Travel Services. I thank him for piquing my interest in forensic neuropsychology, and for providing me with the kind of training in that area that only a true expert could give. I would like to thank the members of my dissertation committee, Kirk Heilbrun, Ph.D., Pam Geller, Ph.D. and Phil Schatz, Ph.D., for their consistent support and constructive feedback, which have helped to shape my thinking about this project.

I am also grateful to my family and friends, who have been there for me through all the disappointment, anxiety, frustration and formatting crises that invariably accompany projects of this magnitude. I am certain that they are pleased that they no longer have to ask me "are you done with school yet?"

Finally, I am grateful to my mom, whose unwavering faith has sustained me through so much.

Table of Contents

List of Tables

List of Figures

Abstract

Trends in Symptom Validity, Memory and Psychological Test Performance as Functions of Time and Malingering Rating

Maryellen McClain

Douglas L. Chute, Ph.D.

Detection of malingered or exaggerated cognitive deficits has received considerable attention within the neuropsychological literature in the recent past. Methods to detect such phenomena have been developed, including specific tests of malingering as well as methods by which to analyze data from existing tests. Questions regarding the utility of these various methods continue to be debated within the field, and neuropsychologists have speculated that litigating patients are being made aware of specific malingering detection methods by attorneys as well as Internet websites devoted to such topics. It was hypothesized that scores on symptom validity measures would normalize over time, with fewer subjects classified as malingering based on their scores on these tests, ostensibly due to prior knowledge of their content and purpose. In addition, significant differences were expected to emerge on tests specific tests of memory using a recognition format as a function of clinical judgment of rating probability. Patients rated as malingerers were expected to exaggerate symptoms of emotional distress with significantly greater frequency than patients rated as not malingering, as evidenced by their responses to the Minnesota Multiphasic Personality Inventory – Second Edition (MMPI-II), the Beck Depression Inventory – Second Edition (BDI-II) and the Beck Anxiety Inventory (BAI), and were expected to make significantly more subjective complaints of memory dysfunction than those patients rated as non-malingerers. Archival data collected on 1290 subjects between 1999 and 2003 were examined in this study, and were extracted from

the Tulane University Neuropsychology Laboratory database. Non-parametric statistical procedures were used, due to the non-normal distribution of the data, and included Kruskal-Wallis analysis of variance procedures and follow-up Mann-Whitney U tests for all pairwise comparisons. The results failed to support the expected trends toward normal scores for the symptom validity measures. However, scores on tests of recognition memory for verbal and visual materials, as well as on the BAI and the Hs and Hy scales of the MMPI-2 were significantly lower for those subjects rated as absolute malingerers compared to those rated as not malingering The rate of subjective complaints of memory dysfunction was similar for all subjects.

1: Background and Literature Review

Prior to the advent of modern neuroradiological tools, neuropsychological assessment provided one of the few means by which organic brain disease and its effects could be diagnosed and described. With the development of more sophisticated and accurate methods for the diagnosis of structural brain pathology, such as computerized tomography (CT) and magnetic resonance imaging (MRI), the role of neuropsychological assessment has evolved to take on more specific diagnostic challenges, such as differentiating between subtypes of disorders, the diagnosis of conditions in which gross structural abnormalities cannot be visualized by existing radiological methods, examining the impact of psychiatric disturbances on cognitive functioning, and determination of disability. The results obtained from such assessments have contributed to ongoing research that has significantly increased our understanding of the brain-behavior relationships with which Neuropsychology is concerned. As a result, more, and more refined, neuropsychological assessment techniques have been developed.

Although many neuropsychologists have typically functioned within a traditional diagnostic and treatment-oriented clinical setting, more and more clinicians are being drawn into the forensic arena. This is especially true regarding cases of personal injury, as an appreciation of Neuropsychology's contributions to the understanding of cognitive functioning after acquired brain injury has grown within the legal community. Within this relatively new domain of practice, the neuropsychologist has been called upon to address such issues as the establishment of causal relationships between the alleged accident and resulting injuries, the impact of premorbid cognitive status on current functioning, and to predict the likelihood that cognitive sequelae of such injuries are either permanent or temporary. Neuropsychological tests are used to establish the severity and permanence of acquired brain injury in such civil litigation cases, in which very large monetary settlements are at often at stake. The neuropsychologist has also been called upon to examine the veracity of patient complaints and to make statements regarding the patient's level of effort and cooperation during the assessment. The methods by which such examinations are accomplished have been the subject of much interest within the domain of forensic neuropsychological practice, and research in this area has grown considerably in the recent past.

According to Brandt (1988), memory dysfunction constitutes the most common complaint after brain injury or neurological illness. Williams (1998) asserts that problems with memory in such circumstances are common knowledge to the general public, via frequent depictions of amnesia and related memory disorders that follow a blow to the head. For that reason, the general public may have a relatively better understanding of the manifestations of memory disorder than of any other cognitive disorders. As a consequence, persons sustaining such injuries may be more likely to feign the presence of memory impairment, even if that injury does not result in an injury to the brain. Alternatively, such patients may also exaggerate real deficits, and the clinician is often called upon to distinguish between these phenomena.

As noted by Iverson and Binder (2000, p. 829), "neuropsychological assessment involves collecting information in the form of symptom reporting and test performance. Both of these types of information can be controlled by a patient who wishes to appear impaired". In an effort to capture such phenomena, the development of tests and assessment techniques to examine symptom validity has become a major endeavor in

Neuropsychology. The success of these methods and tests continues to fall under scrutiny; however, as seen in a professional literature that is fraught with equivocal results and a wide variety of opinion regarding the optimal method for the detection of manufactured or exaggerated complaints. While much of the research published to date has been concerned with the utility of different detection methods in generally well-controlled situations, there have been no studies that have examined the utility of symptom validity tests over time. There have been concerns that litigating patients referred for neuropsychological assessment are being coached, ostensibly by their attorneys, regarding specific tests and malingering detection methods that may be used in the test battery. Also of concern is the growing body of information pertaining to neuropsychological assessment methods that is available on the Internet, some of which presents specific test items as well as tips and strategies to avoid the detection of feigned deficits. As a result, there are concerns that the methods by which malingering and/or symptom exaggeration have been assessed are becoming less effective, due to the wider availability and dissemination of otherwise proprietary test information.

This study examined the utility of some of the methods used to assess symptom validity and malingering in a large sample of litigating and non-litigating patients with a variety of diagnostic backgrounds in an attempt to provide valuable information in this growing area of study.

As noted previously, memory deficits are considered to be the most frequently malingered symptom of cognitive dysfunction (Williams, 1998). Therefore, an understanding of memory, in both normal and impaired subjects, is necessary. The following section provides a brief overview of memory subtypes, and is followed by a discussion of memory functioning in selected populations (“normal” subjects, chronic pain patients, and traumatic brain injury patients). Subsequent sections pertaining to objective tests of memory, discussion of pertinent assessment issues in the context of forensic work, assessment methods designed to detect malingering, and the effects of both psychiatric disturbance and coaching on memory and test performance. Special considerations in malingering research are also discussed.

1.1. Subtypes of Memory

According to Lezak (1995, pg. 27), “central to all cognitive functions and probably to all that is characteristically human in a person’s behavior is the capacity for memory and learning”. Considering the relative importance of memory within the scope of human cognition, it is of little surprise that the capacity to acquire, retain and recall information has been the subject of considerable interest and study for quite some time. For example, Dikmen et al. (1987) note that memory functioning after brain injury has been the subject of investigation more than any other cognitive ability, likely due to its complexity and its importance in daily life and for survival. This undertaking has been difficult; however, as memory has long been appreciated as one of the most complex aspects of brain functioning (Luria, 1981). Memory does not exist as a unitary cognitive construct, nor is it considered to be localized completely to a single, discrete brain structure (Filley, 1995). Comprehensive examination of memory; therefore, must take into account its various facets and examine each of these both individually and in combination. The ability to conduct a thorough assessment of memory functions is of particular importance in the context of brain impairment; as such examinations provide the clinician with the ability to fully delineate the patient's current level of memory functioning, and to make diagnostic and prognostic inferences based on the obtained results.

The assessment of memory generally encompasses the examination of *immediate, short-term,* and *long-term* functions, which describe those abilities that allow for information to be held momentarily, to be held for several minutes to days, or which refer to the storage of memories for very extended periods of time, such as years. References to "learning" generally pertain to the amount of information that is maintained for more immediate use; whereas the term "memory" is applied to information that has been stored and is, theoretically, available for retrieval at a later time (Squire, 1987). Within the domain of immediate memory lies the concept of *working memory*, as described by Baddeley (1992), and which pertains to the patient's ability to not only hold information on a short term basis, but to mentally manipulate that information according to task demands. Theoretically, the working memory model encompasses separate systems that allow for the retention and processing of verbal and non-verbal information, termed the *phonological loop* and the *visuospatial sketchpad*, respectively.

Additional delineations include distinctions between *declarative* memory, which encompasses the patient's recollection of previously presented information, and *procedural* memory, which describes memory for skills and actions. *Semantic* memory includes the patient's knowledge of common facts and general information, whereas *episodic* memory contains information pertaining to specific events. A distinction is also made between *explicit* and *implicit* memory, with the former encompassing conscious recollection as opposed to the unconscious nature of the latter. Finally, amnestic patients are often described in terms of the degree of *retrograde* memory loss, which describes the degree to which memories acquired before the injury have been lost. In contrast, *anterograde* amnesia is concerned with the patient's inability to acquire new information (O'Connor & Morin, 1998). In patients with true amnesia, deficits are generally seen in declarative, episodic and explicit memory, and anterograde amnesia is often present. However, these patients often demonstrate intact procedural, semantic and implicit memory, with variations in the degree of retrograde amnesia. These patterns can be of use to the clinician who is attempting to determine the veracity of a patient's complaints of memory dysfunction.

Despite the ability to delineate subtypes of memory disorder, questions pertaining to the ability of the clinician to determine the validity of the observed memory disorder have received considerable attention. Schacter (1986) proposed that a subjects' *feeling-of-knowing*, or their subjective sense that they would be able to recall a forgotten event if provided specific clues in specific contexts, and this construct proved useful in discriminating real from feigned amnestics when this hypothesis was examined empirically. The results indicated that subjects simulating amnesia tended to minimize their ability to recall specific information, despite the provision of clues, whereas genuine amnestics reported that clues would likely facilitate recall. These results suggested that subjects simulating amnesia did so based on their best guess regarding the nature of true memory loss, which is generally inconsistent with what was expected by these pseudo-amnestics.

1.2. Memory Functioning in Selected Populations

1.2.1. “Normal” Subjects

The degree to which normal subjects endorse cognitive and behavioral symptoms suggestive of central nervous system (CNS) dysfunction has been previously studied (Gouvier, Uddo-Crane & Brown, 1988; Wong, Regennitter & Barrios, 1994), with up to 47 percent of normal subjects endorsing some type of cognitive abnormality. In a sample of medical outpatients with no history of CNS impairment, 62 percent complained of headaches, 26 percent complained of problems with concentration, 20 percent cited memory problems and 16 percent complained of confusion (Lees-Haley & Brown, 1993). These findings suggest that a wide variety of cognitive problems, including complaints of abnormal memory, are endorsed in surprisingly large segments of the general population, presumably in persons without previous head injuries or other forms of neurological impairment.

Patterns of memory test performance have also been useful in distinguishing memory changes resulting from organic impairment from those related to normal aging (Morley, Haxby & Lundgren, 1980). Specifically, dissociation between working memory and memory for information that has already been acquired has been described (Craik, 1977), in which normal elderly subjects are able to hold relatively simple information in working memory, but demonstrate abnormal spontaneous recall that is believed to result from problems with the ability to encode information. Not surprisingly, successful performance on working memory tasks appears to be inversely related to task difficulty, with poorer performances on tasks requiring recall of more complex information. Problems with acquisition have been correlated with a decline in cognitive functions in general, resulting in less efficient strategies for encoding material that must be remembered (Sanders et al., 1980). Problems with retrieval also appear to be related to this generalized decrease in cognitive efficiency, resulting in less efficient retrieval of information and a reduced ability to utilize cues to facilitate recall (Craik, 1977).

Palmer et al. (1998) examined the degree to which cognitive problems occur in normal elderly subjects. Using a comprehensive, flexible neuropsychological test battery, 73 percent of their group of 132 older (aged 50-79) subjects performed within the Borderline Range on at least one cognitive test, while 37 percent of the sample generated at least one score in the Impaired Range. However, only five percent of the sample generated scores that were consistently within the Impaired Range. Within the domain of memory, only 0.8 percent of the sample was consistently within the Borderline Range and none was consistently within the Impaired Range. Therefore, the authors caution that false positive interpretations of cognitive impairment are possible in otherwise normal subjects, without consideration of these, and similar, base rates.

1.2.2. Chronic Pain

Patients diagnosed with chronic pain have been found to make subjective complaints of cognitive dysfunction, with approximately 42 percent of patients making such complaints in one sample (Iverson & McCracken, 1997). Additional analysis of this sample revealed that 39 percent of the patients met criteria for a diagnosis of Postconcussive Disorder, according to the diagnostic criteria set forth in the DSM-IV (American Psychiatric Association, 1994), even though none of the patients had any history of head injury. These findings suggest that complaints of memory dysfunction are often non-specific.

Iverson et al. (2001) examined the rate of cognitive complaints in litigating and non-litigating chronic pain patients, as well as patients with acquired brain injury. In their sample, the non-litigating chronic pain patients comprised the group making the fewest complaints of cognitive dysfunction, whereas the pain patients seeking Workman's Compensation payments complained of more cognitive problems than a group of non-litigating patients with documented brain injuries. The authors caution; however, that their results may have been affected by the fact that the non-litigating pain patients were all candidates for surgical implantation of dorsal column stimulators and, in the context of pre-surgical psychological evaluations, may have denied any form of dysfunction so as to not jeopardize their chances of receiving this potentially pain-reducing device. Conversely, the patients involved in litigation or Workman's Compensation proceedings may have perceived the need to appear impaired in order to justify their requests for financial compensation. These findings raise an important issue within the domain of self-report of cognitive impairment, namely, that such complaints may often be spurious and motivated by external factors.

1.2.3. Acquired Brain Injury

Although patients can suffer from any number of problems associated with brain injuries, those that occur within the cognitive and neurobehavioral domains are considered to account for the majority of the deficits that endure over time (Brooks, 1986). In particular, memory functioning after acquired traumatic brain injury has been the subject of considerable research, with many studies concerned with the extent and course of recovery of memory as a function of injury severity. The Glasgow Coma Scale (GCS; Teasdale & Jennett, 1974) is generally used to evaluate and objectify the patient's level of consciousness, and is considered by some to be a useful assessment of injury severity in patients with altered consciousness (Lezak, 1995, p. 755). The scale provides criteria for the assessment of the patient's ability to open their eyes, and to respond verbally and motorically, according to the level of input of stimuli from the examiner (from verbal commands to infliction of pain). Scores are generated from this assessment, with scores equal to or above 13 indicative of coma duration of less than or equal to 20 minutes, corresponding to a mild level of brain injury. Moderately severe brain injuries manifest coma duration of 21 minutes to six hours, and scores of 9 to 12 on the GCS. Those injuries classified as severe demonstrate coma duration of longer than six hours and GCS scores of 8 down to 3. The GCS can be administered on a repeated basis in order to track the depth and length of coma, with longer and deeper comatose states associated with more severe brain injuries (Dikmen, et al., 1986a; Dikmen et al., 1990.).

While the ability to determine the severity of brain injury is of considerable significance, the ability to determine the degree to which patients recover cognitive functioning after such an injury is also important. In that regard, the base rates of cognitive impairment after brain injury must be considered, in order to provide a basis against which to compare neuropsychological test data from an individual patient. Failure to do so often results in over or under diagnosis of organic impairment (Larrabee, 2000) and erroneous conclusions regarding the etiology of the patient's cognitive problems. In a comprehensive study of 436 brain injured adults, Dikmen et al. (1995) demonstrated that, at one year post-injury, their sample of severely brain -injured patients was significantly more impaired on all cognitive measures, including memory tests, than patients for whom mild brain injuries had been documented. Moreover, those patients whose injuries were within the mild range, according to their initial rating on the GCS, generated scores on cognitive tests, including memory measures, that were not significantly different from the control group, which was comprised of patients who had sustained trauma to body parts other than the head. In fact, the authors state, "significant neuropsychological impairment due to a mild head injury is as unlikely as is escaping impairment in the case of a very severe head injury" (Dikmen et al., p. 87). These findings were consistent with other studies that suggested that mild brain injuries, defined as those in which the patient's GCS score ranged from 13 to the maximum of 15, are not associated with persistent neuropsychological deficits (Dikmen, McLean & Temkin, 1986b; Levin, et al., 1987; Maddocks & Saling, 1996). While Brandt (1988) notes that head injury patients complain most frequently of problems with memory, it has also been demonstrated that moderately to severely injured brain injury patients tend to underestimate the severity of their deficits, due mainly to decreased self-awareness (Sherer et al., 1998a; Sherer et al., 1998b) and may; therefore, report fewer problems than those with less severe, or no, brain impairment. This pattern of self-report may be useful in discriminating patients with such injuries from those who falsely report such symptoms, or exaggerate milder cognitive problems.

While the presence of cognitive impairments, such as slowed speed of information processing, attention, and memory, have been documented in mild brain injury patients assessed up to one week post injury (Maddocks & Saling, 1996; Levin et al., 1987), other studies have demonstrate that, in up to 96 percent of such patients, these problems are largely resolved within approximately three months of the injury (Dikmen et al., 1986a; Binder, Rohling & Larrabee, 1997). The base rate of persistent cognitive deficits in patients sustaining mild brain injuries has been estimated to be approximately five percent, suggesting that it would be very likely that a diagnosis of persistent deficit (beyond three months) that is directly attributable to the injury would be incorrect (Binder, Rohling & Larrabee, 1997). Despite these findings, it has also been noted that a subset of mild brain injury patients complain of persistent cognitive impairment and, according to Mittenberg and Strauman (2000), these patients are more likely to pursue litigation than the majority of mild brain injury patients. The descriptive classification of Postconcussion Syndrome (hereafter referred to as PCS) is often applied to such patients. As defined by the *International Statistical Classification of Diseases and Related Health Problems,* 10th Edition (World Health Organization, 1992), PCS requites the presence of problems in three or more categories including emotional disturbance, subjective complaint of cognitive problems despite a lack of such findings on formal assessment; headache, noise intolerance and vestibular disturbance; insomnia and alcohol intolerance. A preoccupation with these problems, suggestive of Hypochondriasis, is also part of the disorder and is accompanied by a fear of permanent brain damage and adoption of the sick role. The presence of depression and anxiety, borne out of fear of the possible permanence of the cognitive dysfunction, is considered to perpetuate the patient’s problems. The patient’s involvement in litigation is not; however, associated completely with the presence of the syndrome, according to the diagnostic criteria. The *Diagnostic and Statistical Manual of Mental Disorders – IV* (APA, 1994) also provides criteria for a diagnosis of PCS. These criteria require a history of head trauma resulting in a loss of consciousness, posttraumatic amnesia, or seizures, evidence of abnormalities in attention or memory as documented on neuropsychological testing. The DSM-IV and also requires that three or more symptoms either begin shortly after the injury or are an exacerbation of previous symptoms, and which persist for at least three months. These symptoms include those described in the ICD-10 criteria, but also include personality changes, apathy and lack of spontaneity.

The validity of the PCS diagnosis has been the subject of ongoing debate in neuropsychology, with some clinicians maintaining that the homogeneity of PCS symptoms suggests an organic etiology (Binder, 1996), while others implicate a purely psychogenic basis (Bohnen, Jolles & Twinjstra, 1992; Youngjohn, Burrows & Erdal, 1995), and still others state that while the initial problems may result from neurological insult, the patient's perception of persistent cognitive problems is psychologically-based (Mittenberg & Strauman, 2000). It has been demonstrated that patients who have sustained some form of non-physical trauma (non-brain injuries or emotional trauma) report similar rates of PCS symptoms when compared to mild brain injury patients (Lees-Haley, Fox & Courtney, 2001). Furthermore, the presence of PCS-like symptoms has been found to vary on a daily basis in non-brain injured subjects, in relation to the subject's perceived stress level (Gouvier, et al., 1992). The presence of PCS symptoms has also been documented in the general population (Gouvier, Uddo-Crane & Brown, 1988; Mittenberg, Tremont & Zielinski, 1996; Lees-Haley & Brown, 1993).

Information regarding the patient's subjective perception of their cognitive problems is also important in neuropsychological evaluation, as it provides additional information to which the test data can be compared. However, the validity of the patient's self-report of such problems is often of questionable validity. Nowhere is this more evident than in the forensic setting, in which many patients undergo neuropsychological evaluation in an effort to document the presence of problems that, if validated and deemed permanent, may be deserving of significant financial compensation. Using a self-report checklist of 37 possible problems, Lees-Haley and Brown (1993) found that personal injury claimants endorsed significantly more symptoms of PCS and other problems than a control group of general medical patients who were not involved in litigation. It is interesting to note that, in this litigating sample, the most commonly reported problem was anxiety or nervousness (93 percent of the sample), while sleep disturbance, depression and headaches were also endorsed by more than 80 percent of these respondents. Only 53 percent of the litigating sample, which was claiming significant cognitive impairment due to head trauma, reported memory problems; this was considered to be a relatively low number given that memory problems constitute the most frequently reported sequelae of acquired brain injury, as noted previously (Brandt, 1988). Additionally, Fee & Rutherford, 1988 demonstrated that, while litigating and non-litigating patients reported similar rates of PCS symptoms at six weeks after injury, the rate of reported symptoms in non-litigating subjects was significantly lower than their litigating peers, even when this latter group reached settlement of their lawsuits. Several authors (Trueblood & Schmidt, 1993; Trueblood, 1994; Binder et al., 1993) have found that litigating patients who perform below acceptable levels on malingering tests also tend to generate lower memory test scores than patients whose brain injuries have been verified and who are not involved in litigation. In addition, their memory test performances tend to fall below those of litigating patients who perform within acceptable limits on tests of malingering. Based on such findings, Lees-Haley and Brown (1993) caution that patient self-report of cognitive dysfunction must be viewed with a skeptical eye, especially within the context of a forensic neuropsychological evaluation.

1.3. Neuropsychological Assessment of Memory

Given that memory dysfunction “is commonly reported in individuals diagnosed with a wide variety of neurological, psychiatric and developmental disorders” (Wechsler, 1997, p. 1), the need for methods to comprehensively assess memory is of utmost importance. To meet that need, a number of test instruments specific to the assessment of memory have been developed, with revised versions of some of these tests now available that attempt to incorporate additional methods by which to assess memory. A brief summary of some of the more commonly used instruments is provided here.

1.3.1. Wechsler Memory Scale

The revised version of the Wechsler Memory Scale (WMS-R) was introduced in 1987, and was described as “an individually administered, clinical instrument for appraising major dimensions of memory functions in adolescents and adults” (Wechsler, 1987, p. 1). Although presented as a single instrument, the WMS-R is essentially a battery of individual tests designed to examine multiple aspects of memory functioning, and represents a significant expansion and refinement of the original Wechsler Memory Scale (Wechsler, 1945). The WMS-R allows the clinician to differentially examine the patient’s immediate memory functioning within the auditory and visual domains, and also provides indices of general memory, delayed memory and attention/concentration. Normative studies conducted with the WMS-R in a sample of 20 brain-injured patients demonstrated that, as a group, these patients performed at generally lower levels compared to normal subjects. In particular, brain injured subjects generated Delayed Memory Index scores that were significantly lower than their General Memory Index scores, suggesting relatively greater problems with recall after a time delay. In addition, the Delayed Memory Index for the brain-injured sample was significantly lower than the mean Attention/Concentration Index for this group, further illustrating problems with learning and spontaneous recall. A substantial revision of the WMS-R was completed in 1997, resulting in the introduction of the Wechsler Memory Scale – Third Edition (WMS-III). According to the manual (Wechsler, 1997, p. 1), the WMS-III is "an individually administered battery of learning, memory, and working memory measures". This description highlights the changes in the conceptualization of memory functioning that occurred in the 10 years between the publication of these two instruments, including the emphasis on the concept of working memory as a measure of information processing capacity. Also, subtests designed to examine differences in performance on recall versus recognition measures were added to allow the clinician to more fully examine specific problems with the retrieval of information that would, hopefully, aid in the differential diagnosis of subtypes of memory disorders. Results of the normative studies conducted with brain injured subjects indicated that, as a group, these patients demonstrated mild problems with encoding and storage of information, especially within the visual domain, but were average with respect to the ability to discriminate target information from foils on recognition-based subtests. This latter finding is of particular interest, as the use of recognition memory tests has received increasing attention in the symptom validity literature. An appreciation of typical patterns of performance on such tests in both normal and neurologically impaired subjects is needed, in order to provide a standard against which the recognition memory test performances of subjects believed to be malingering can be compared.

Like its immediate predecessor, the WMS-III is a comprehensive battery of tests designed to measure immediate and delayed memory in both the verbal and visual domains. Eight index scores are calculated from this test data, including the Auditory Recognition Delayed Index (ARD), which provides a measure of the patient’s ability to discriminate target verbal information from foils, and represents the combination of two of the verbal subtests found on the WMS-III (Logical Memory-II Recognition and Verbal Paired Associates-II Recognition). The Recognition portion of the Visual Reproduction-II subtest is used to examine the patient’s recognition memory for non-verbal materials; in this case, a set of five geometric designs of increasing complexity. These scores (ARD and the Visual Reproduction-II Recognition subtest) are of particular interest in this study, as both provide measures of recognition memory.

1.3.2. Rey Auditory Verbal Learning Test

This list learning task (RAVLT; Rey, 1964) was designed to measure immediate memory span, learning over repeated trials, the presence or absence of learning strategies, and the incidence and effects of proactive and/or retroactive interference in addition to the degree of short-term and long-term recall. The target list consists of 15 unrelated words that are read to the patient over five consecutive learning trials, followed by an interference trial also consisting of 15 words, a sixth trial testing recall of the target list, a delayed recall trial of target list items administered 30 minutes after initial presentation, and a recognition trial testing the patient's ability to discriminate target list items from foils.

Normative studies using the RAVLT indicate that normal subjects typically demonstrate increased acquisition of the target list over the five learning trials (Selnes et al., 1991), with recall on the interference trial falling an average of one to two words below recall on trial V, and approximately the same amount of loss between trial V and VI (Lezak, 1995). Lezak (1995, p. 442) notes, "marked variations from this pattern will usually reflect some dysfunction of the memory system". The use of this test with brain-injured patients has demonstrated that, in general, such patients demonstrate some acquisition over the five learning trials with poor delayed recall, but are generally within normal limits in their performance on the recognition memory trial (Bigler et al., 1989).

1.3.3. California Verbal Learning Test

The California Verbal Learning Test (CVLT; Delis, et al., 1987) was developed to provide "a brief, individually administered assessment of multiple strategies and processes involved in learning and remembering verbal material" (Delis, et al, 1987, p. 1). The test presents the patient with a list of 16 words from four different semantic categories (i.e. spices, articles of clothing, tools and fruits) over five consecutive learning trials. Recall is tested immediately following each presentation of the list. An interference trial, consisting of a different list of 16 semantically related words, is administered immediately after completion of the initial five learning trials. Measures of free and cued recall are administered immediately after completion of the interference trial, and again after a 20-minute delay. A recognition memory trial is also administered after completion of the cued delayed recall trial. Additional scores pertaining to the subject's ability to discriminate target words from foils (Discriminability) and their tendency to respond either affirmatively or negatively on the recognition trial (Response Bias) are also calculated.

The CVLT, in its original form, provides a wealth of information regarding specific aspects of a patient's ability to acquire, retain and recall verbal information. Within the province of malingering assessment, particular attention has been paid to differences between spontaneous recall and recognition of the list elements. As noted in the test manual (p. 27), "recognition testing maximally aids retrieval, whereas cued-recall testing provides a lesser degree of assistance. In contrast, free recall provides no assistance". Therefore, it would be expected that those patients experiencing problems with retrieval would benefit from information provided in a recognition format. Those patients who cannot encode information (amnestics) would be expected to perform poorly on both recall and recognition trials. Both scenarios provide guidelines for expected patterns of performance on this test. Research using the CVLT has been conducted with a multitude of special populations in order to examine such patterns. Crosson et al. (1988) demonstrated that severely brain injured patients were able to generate learning curves similar to normal controls, but acquired less information per trial than their normal control counterparts. The brain injured patients also recalled significantly fewer total words than normals on delayed recall, and also performed below normals on the recognition trial; however, the brain injured patients' scores demonstrated that they derived benefit from the provision of cues on this trial. No differences were found; however, between patient and controls with respect to response bias, defined as the tendency to respond with either “yes” or “no” to individual test items on the recognition trial. Novack et al. found similar effects (1995) in another sample of severely brain-injured patients. Depressed patients have also been found to score below normal levels on all CVLT indices, but their performance tends to improve on the recognition trial (Massman et al., 1992).

A revision of the CVLT, the CVLT-II, was completed in 2000, and incorporated several changes to the original test format, including a larger sample (1087 subjects compared to 273 for the CVLT). Additional changes include word lists that were comprised of more familiar words, a short form for more severely impaired patients with limited tolerance for testing, and the addition of a forced-choice recognition trial to aid in the assessment of suboptimal effort or malingering. The CVLT-II, as noted previously, is used as a measure of the patient’s ability to acquire, retain, and recall verbal information both immediately and after a time delay, in much the same manner as the RAVLT and the original CVLT. However, the word lists used in the CVLT and CVLT-II can be organized into four semantically related lists (i.e. vegetables, tools, items of clothing, modes of transportation) that, theoretically, provide a structure that facilitates recall of individual list items. The revised version of the CVLT is distinguished from the original by the inclusion of several new subtests, including a yes/no recognition subtest examining recognition memory for List A (target) items. In this subtest, which is administered at the end of the test, the patient is read a list of 16 word pairs, with each pair containing a target word from List A and a foil. The patient is required to identify the word that was presented in List A, and the patient's performance is expressed as the percent of List A words that are identified correctly. The words in each pair on this recognition trial are further distinguished by their level of complexity, with the List A words being relatively concrete nouns (i.e. "onion", "cow", "desk") whereas the foils are more abstract nouns that are encountered with less frequency in everyday discourse. (i.e. "majority", "technique", "sprinkler"). Connor et al. (as cited in Delis, Kramer, Kaplan & Ober, 2000) tested this format in a population of traumatically brain-injured patients, as well as a population of subjects exhibiting suboptimal effort on the Hiscock forced choice test. The results revealed that this forced choice format had 80 percent sensitivity and 97 percent specificity in correctly identifying patients attempting to feign memory deficits. The CVLT-II normative data indicate that more than 90 percent of the entire normative sample scored perfectly on this forced choice subtest (Delis, et al., 2000, p. 162). These very high ceiling effects demonstrate the utility of this subtest as a measure of effort, as the overwhelming majority of patients, even those with significant cognitive impairments, are able to complete this subtest without error. Given this finding and its pertinence to the study of malingering and dissimulation, subject performance on the forced-choice trial was of primary interest in this study.

1.4. Neuropsychological Assessment in the Forensic Setting

The aims of neuropsychological assessment within the forensic arena differ with respect to the types of patients being evaluated, and are generally concerned with the establishing the presence of acquired brain injury and, if such is verified, describing the extent and likely permanence of the resulting cognitive impairments. According to

Kolpan (1996, p. 102), “the measurement of the claimant’s loss is the disparity between the functional abilities before the trauma and his or her disability after sustaining a TBI”. Generally, in forensic cases concerning moderate to severe brain injury, where the injuries are generally well documented, the neuropsychological exam is focused on the determination of the type and severity of the patient’s cognitive deficits and the degree to which they are related to the events in question. However, the examination of a patient with mild brain injury is often focused on establishing the probable presence or absence of acquired brain impairment (Sherer, Madison & Hannay, 2000). In many cases, when questions pertaining to the possibility of permanent injury and disability are under consideration, the incentive to malinger, or to intentionally perform below one’s true cognitive capabilities, can be quite attractive to the litigating patient, considering the monetary awards that are often at stake. Therefore, methods to accurately detect such manufactured performances are an essential part of the forensic neuropsychological evaluation. Complicating this issue; however, is the possibility that a patient’s less-than-optimal performance is due to factors other than a conscious attempt to appear cognitively impaired. These factors can include a variety of psychiatric disorders, such as mood, Factitious and Somatoform disorders. Therefore, all possible explanations for seemingly implausible performances must be investigated during the course of the forensic neuropsychological evaluation, and making a diagnosis of malingering, or some other disorder, requires the clinician to consider the various diagnostic criteria that are required in each case. These criteria are discussed below.

1.4.1. Diagnostic Criteria for Malingering

The ability to differentiate intentional from unintentional symptom production and symptom exaggeration is of considerable importance, as the personal damage that can result from mislabeling a patient as a "malingerer" can be devastating and difficult to eradicate from one's record. Conversely, one would presume that the courts and insurance companies would prefer to avoid awarding large sums of monetary damages to otherwise unimpaired individuals who have been able to feign permanent cognitive impairment in a sufficiently convincing manner. Therefore, careful consideration of the methods used to evaluate and confirm such diagnoses, and of the information these methods provide, is of utmost importance.

The term "malingering" is often used indiscriminately to describe patients whose test data do not completely adhere to the patterns of performance expected for their presenting problem. Clarification of the term "malingering", therefore, seems warranted. The *Diagnostic and Statistical Manual of Mental Disorders,* Fourth Edition (DSM-IV; 1994, pg. 683) defines malingering as the "intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by external incentives". This definition is unchanged in the DSM-IV Text Revision (DSM-IV-TR; 2000, p. 739). However, as noted by Suhr et al. (1997, p. 500) "one cannot simply assume that impaired test performance in a forensic context is reflective of malingering". The emphasis on the intentional nature of the DSM-IV definition is particularly important, as it separates malingering from other sources of symptom production or exaggeration where the etiology of the behavior is less well defined, such as Somatization, Conversion or Factitious disorders, suggestibility, or mood disorders.

The base rates of malingering must also be considered, with the additional understanding that the incidence of such behavior appears to vary widely in different clinical groups. Solid data on this subject are lacking, and the available data vary with respect to the environment in which they were collected. In survey research conducted by Rogers, Sewell and Goldstein (1994), a sample of 320 forensic psychologists in mental health settings estimated that approximately 16 percent of forensic and 7 percent of non-forensic patients attempt to malinger. A literature review conducted by Rogers, Harrell and Liff (1993) suggested that up to 50 percent of patients undergoing neuropsychological evaluation in the context of litigation might be malingering. Although additional research regarding the establishment of base rates of malingering is needed, it is also compounded by the methodological problems inherent in identifying a subject group that is, by definition, intent on escaping detection. It is also necessary to differentiate malingering from other sources of symptom production or exaggeration, as noted previously; therefore, familiarity with the different diagnostic criteria for each category is required, and these are discussed in the following section.

1.4.2. Differential Diagnosis in the Assessment of Malingering

As noted above, other phenomena must be considered, investigated and ruled out before the label of "malingerer" is applied to a patient. In most cases, alternative explanations for unusual test scores and behavior can be found in other DSM-IV diagnostic categories, such as the Somatoform, Mood and Factitious disorders (Iverson & Binder, 2000). Factitious disorders are closely related to malingering, in that both involve the intentional production of symptoms. The difference between these conditions;

however, lies in the reward that is sought by the patient. As noted above, malingering patients are motivated by external rewards such as money; whereas, the patient with a factitious disorder assumes or perpetuates a sick role in order to obtain some internally reinforcing reward (i.e. eliciting and gaining attention and caretaking from another individual).

The Somatoform Disorders, according to the DSM-IV (1994, p. 445), comprise those disorders that present with symptoms suggestive of physical ailments, but for which no underlying disease state can be identified. Each of these diagnoses is distinguished from Factitious Disorders and Malingering in that the symptoms are not consciously produced by the patient. A significant psychological component is generally identified in each of these disorders; however, distinctions can be made between the various diagnostic categories. In Somatization Disorder (APA, 1994, p. 446), a constellation of symptoms that are manifested before the age of 30 are required for the diagnosis to be made, including pain, gastrointestinal problems, sexual problems, and pseudoneurological problems. In Undifferentiated Somatoform Disorder (APA, 1994, p. 450), the patient presents one or more physical complaints (most commonly fatigue, appetite loss, and gastrointestinal or genitourinary problems) for at least six months. In contrast, patients who meet criteria for a diagnosis of Hypochondriasis (APA, 1994, p. 462) do not present with a specified group of symptoms, but make ongoing complaints of medical illness or general physical malfunctioning despite evidence to the contrary. Patients diagnosed with Conversion Disorder (APA, 1994, p.452) frequently present with pseudoneurological problems, such as loss of sensation or movement, for which no underlying neurological problem can be identified. In contrast to the above disorders, Pain Disorder (APA, 1994, p. 458) encompasses an actual physical experience of pain, but maintains that psychological factors are central to the onset, severity and continuation of the pain.

1.4.3. Assessment of Mood and Emotional Functioning

Patients with Mood disorders may also generate patterns of test performances that raise questions regarding their validity. While hallmark features of Major Depressive Disorder (APA, 1994, p. 320) involve significant emotional problems such as sadness and hopelessness, additional features are found within the realm of cognitive functioning and include: difficulty with information processing, poor attention and concentration, reduced decision-making abilities, and impaired memory. Additionally, behavioral manifestations of Major Depression can include irritability, poor or inconsistent cooperation, and psychomotor slowing. An irrationally negative self-perception is also frequently seen in depressed individuals, resulting in abnormal and inaccurate self-appraisal which, in turn, can lead to a distorted and inaccurate self-report of one's abilities. This constellation of problems can confound attempts to arrive at a valid diagnosis for a given patient, as these symptoms overlap considerably with those seen in neurologically impaired patients, as well as those attempting to malinger or exaggerate their symptoms.

Cronholm and Ottonsson (1961) explored cognitive functioning in a sample of depressed patients and found that these patients demonstrated deficits on objective measures of memory. More specifically, their patients scored abnormally on tests of immediate and delayed spontaneous recall, suggesting impaired learning; however their performances on measures of recognition memory were generally within normal limits. In general, the results suggested problems with spontaneous recall, rather than a deficit in the acquisition of target information. Despite these findings, the utility of neuropsychological testing in the differentiation between organically and psychiatrically based cognitive impairment has been questioned (Smith et al, 1976; Nott & Fleminger, 1975). In fact, in 1986, Caine stated “we have no sound method for demarcating the boundaries between (1) intellectually intact depressed elderly individuals; (2) others who have significant affective symptoms and substantial cognitive impairment…and (3) those who suffer a progressive neurological disease which manifests itself with both behavioral symptoms”. Despite this view, Wells (1979) found that inconsistent patterns of test performance were routinely produced by his group of psychiatric patients, in contrast to the more consistent test score patterns demonstrated by patients with documented organic dysfunction. He interpreted this inconsistency to be indicative of the variable nature of depression and its subsequent impact on neuropsychological test performance, and considered such test score scatter to be a hallmark of mood disturbance. Additionally, depressed persons’ problems with attention tend to interfere with the amount of effort expended in any task, including the acquisition and processing of new information (Caine, 1986). Accordingly, problems with inefficient acquisition strategies are associated with poor recall, as demonstrated by Weingartner et al. (1981). When information was presented in a structured format, depressed patients performed similarly to controls. However, with successive decreases in the structure with which target information was presented, the performances of depressed patients declined on such tasks.

It has long been understood that patients who sustain significant brain injury often develop affective disturbance, associated with the loss of premorbid functioning and independence (Lezak, 1995; Fann et al., 1995; Reitan & Wolfson, 1997). However, studies examining the impact of emotional state on cognitive functioning in these patients, especially memory, have returned equivocal results (Gass, 1996; Alfano et al., 1993), due in large part to differences in methodology. Despite the lack of a consensus on this problem in the brain injured population, relatively more consistent data have been found regarding the association between emotional functioning and memory in both normals and psychiatric patients. Several authors have found such an association between depressive disorders (Zakzanis, Leach & Kaplan, 1998; Kindermann & Brown, 1997), schizophrenia (Kareken, Moberg & Gur, 1996) and memory. In contrast, Kizilbash, Vanderploeg & Curtiss (2002) demonstrated that depression alone did not exert a significant detrimental effect on performance on the CVLT in a group of 399 veterans; however, patients with co-morbid depression and anxiety performed significantly worse on measures of both immediate and delayed recall, as well as the overall amount of information acquired in the course of this test.

Although these findings are illustrative, any mechanism that might explain such an association between psychiatric state and memory has yet to be identified. Some work has suggested that emotional problems exert significant adverse effects on tasks requiring effortful processing by increasing demands for attention, which is vital to the ability to perform optimally on memory tasks (King & Caine, 1996; Zakzanis, Leach & Kaplan, 1998). An association between emotional state, attention and memory was observed by Adams et al. (2001), in that attention acted as a mediator between affective state and memory. These authors interpreted their results to suggest that emotional problems prevent the patient from focusing their full attention on the information to be acquired, thereby interfering with the patient's ability to encode such information at expected levels.

Some research has been devoted to examining the degree to which mood disturbance can exacerbate, or account for, complaints of neurobehavioral and/or cognitive dysfunction. Iverson and Binder (2000) caution that the cognitive problems associated with depression, including slowed information processing speed, reduced concentration, distractibility, and reduced memory functioning, can interfere with performance on both cognitive and symptom validity tests, leading to erroneous conclusions regarding the validity of a patient's performance. They caution that depression is "an important differential when considering the possibility of malingering or other diagnoses" (pg. 833). In a related study, Trahan, Ross and Trahan (2001) found that their sample of depressed subjects endorsed significantly more post-concussional symptoms than a group of mild brain injury patients. Despite the observations noted above regarding the similarity between acquired cognitive problems and those experienced by depressed patients, the depressed sample in this study had no prior diagnosis of brain injury or any other medical problems that would better account for their cognitive dysfunction, yet this sample endorsed symptoms typically associated with post-concussion syndrome and not with clinical depression (i.e. getting lost, slurred speech, chest pains).

The foregoing discussion suggests that interest in the association between mood and cognitive functioning has grown, and has begun to expand into more specific areas of study, including the degree to which mood state accounts for performance on symptom validity tests. Despite growing interest in the possible association between mood and malingering, relatively little research on this topic has been undertaken to date, although some studies have revealed interesting findings. For example, Suhr et al. (1997) found that suspected malingerers generated higher scores on the Beck Depression Inventory (BDI; Beck, 19) relative to patients diagnosed with somatoform disorders. McClain & Black (2002) found that anxiety, as measured by the Beck Anxiety Inventory (BAI; Beck, 1990) accounted for a significant portion of the variance on all three trials of the Word Memory Test (WMT) (Green, 1996). In the same study, performance on the Beck Depression Inventory – Second Edition (BDI-II; Beck, 1996) was predictive of performance on all three trials of the Test of Memory Malingering (TOMM; Tombaugh, 1996). Contradictory findings; however, have also been published. Rees, Tombaugh & Boulay (2001) demonstrated that a sample of inpatients diagnosed with Major Depressive Disorder performed in a manner similar to controls on all three trials of the TOMM, suggesting that this test is resistant to the effects of major psychiatric illness. The limited research conducted regarding the association between mood and performance on symptom validity tests, and the equivocal results found in many of these studies, suggests that ongoing investigation is warranted in this area.

Patients who malinger often produce distinct patterns of performance on one or more neuropsychological tests, which can vary depending on the nature of the examination. For example, patients wishing to appear impaired may exaggerate the loss of either cognitive or physical functioning (Larrabee, 1998), or both (Greiffenstein, Goal & Baker, 1995; Youngjohn, Burrows & Erdal, 1995). One method of examining the degree to which patients exaggerate such complaints involves an examination of Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) and Minnesota Multiphasic Personality Inventory – Second Edition (MMPI-2; Hathaway & McKinley, 1989) test profiles, as many of the test items are specific to physical and cognitive functioning, and often question the presence of unusual or relatively infrequent phenomena within these domains. The MMPI and MMPI-2 are self-report inventories of psychological functioning designed to assist the clinician with the diagnosis of a number of mental disorders. Built into these inventories are scales designed to assess the veracity of the respondent's answers, based on the likelihood that a given question should be answered "true" or "false", even in a clinical sample (i.e. responding "false" to the question "I do not always tell the truth"). Analysis of test profiles, with particular attention to those scales that assess the overall validity of the subject's responses, has become common practice in forensic neuropsychological assessment. According to Iverson and Binder (2000), patients attempting to malinger neuropsychological impairment do not typically produce invalid MMPI-2 profiles, which are often characterized by elevations in the F (Infrequency) scale, which indicate the endorsement of highly unlikely symptoms and problems. On the contrary, such patients frequently generate valid profiles with significant elevations on specific clinical scales, such as 1 (Hs or Hypochondriasis), 2 (D or Depression) and 3 (Hy or Hysteria) (Youngjohn, Burrows & Erdal, 1995; Larrabee, 1998; Suhr et al., 1997). The Hs or Hypochondriasis scale was designed to assess a given subject's "abnormal, psychoneurotic concern over bodily health" (Butcher & Williams, 1992, p. 63). The D or Depression scale assesses depressive symptomatology, and the Hy or Hysteria scale assesses symptoms of conversion disorder, manifested in the development of significant physical disability in the absence of identifiable medical disease. Using the MMPI, Heaton et al. (1978) reported that subjects instructed to feign head injury symptoms generated test profiles with significant elevations on the F scale, as well as scales 1 (Hypochondriasis), 3(Hysteria), 6 (Paranoia) and 7 (Psychasthenia). Moreover, the observed elevations were significantly higher than those generated by patients for whom moderate to severe brain injuries had been documented. Similar response patterns have been found using the MMPI-2, in which groups of litigating head injured patients and simulating malingerers have generated significant elevations on scales 1 (Hypochondriasis), 2 (Depression), 3 (Hysteria), 7 (Psychasthenia) and 8 (Schizophrenia) (Berry et al., 1995). Hoffman, et al. (1999) found elevations on these scales in litigating TBI patients, irrespective of the severity of their injuries; but also found that, within the same study, elevations on these scales were inversely related to head-injury severity in patients who were not litigating. The use of the MMPI to investigate the presence of psychopathology in closed-head injury patients has been met with some criticism, due to the notable similarities between symptoms of psychiatric and organic disturbances. To that end, some investigators have suggested the use of a “neurocorrection factor” to account for scale elevations that occur as a function of head-injury sequelae rather than true psychopathology (Alfano, Paniak & Finlayson, 1993). Other studies have failed to find support for the use of this factor (Dunn and Lees-Haley, 1995), and have suggested that test items that comprise the factor may be more sensitive to the effects of depression than brain injury (Brulot, Strauss & Spellacy, 1997).

Lees-Haley, English and Glenn (1991) noted that litigating patients attempting to malinger often generated MMPI-2 test profiles in which they appeared to combine simultaneous attempts to both deny and exaggerate symptoms of emotional distress on the same test profile. The authors suggested that such patients attempted to portray themselves as honest, well-adjusted, high-functioning individuals who were experiencing acute trauma and permanent disability. In an attempt to provide a method by which to examine such a mixed response set, these authors developed the Fake Bad Scale (hereafter referred to as the FBS), which is comprised of 43 responses that purport to discriminate such patients from those responding in a more straightforward and honest manner. In the 1991 study, this scale was reported to correctly classify up to 93 percent of simulated malingerers, and 96 percent of litigants who were suspected of malingering. A false positive rate of 10 percent; however, was reported. Other research has demonstrated a correlation between the FBS and total correct and response latency measures on the Victoria Symptom Validity Test (Slick, et al., 1996). In addition, the FBS has demonstrated an ability to detect patients attempting to malinger somatic complaints and who also produced cognitive test score patterns suggestive of malingering, as well as MMPI-2 elevations on scales 1 and 3 (Larrabee, 1998). Iverson and Binder (2000, p. 852) state, “the FBS appears to be a promising MMPI-2 scale for identifying individuals who may be exaggerating symptoms”, but caution that “additional research on the sensitivity, specificity and predictive power of the scale will facilitate clinical use”.

Lees-Haley (1997) has also examined the base rates of elevations on the validity scales, basic clinical scales and PTSD scales in a group of 492 patients involved in personal injury litigation. The PTSD scales (PK and PS) were included in the analyses as they were potentially applicable to the subjects in the sample, all of who had allegedly experienced some form of physical or emotional trauma (i.e. brain and/or spinal cord injury, burns, musculoskeletal injury, sexual harassment, wrongful termination, etc.). The results indicated that, for men and women, the most common MMPI-2 codetype was 1-3/3-1, which was produced by 41 percent of the sample. For men, the 1-2/2-1 and 2-3/3-2 codetypes were the second and third most common, respectively. However, this order was reversed for the female subjects. These results differ considerably from those of the MMPI-2 normative sample, in which scales 1 and 3 were the least frequently elevated for men, and were the second and third least elevated, respectively, for women. With respect to the validity indicators in the Lees-Haley study, 23 percent of the sample produced an F score that exceeded a T score of 70, 19 percent had an F-K index greater than zero, 22 percent had a K index greater than a T score of 40, 33 percent had an L index below 49, and 20 percent had an Fb index that exceeded a T score of 70. For the PTSD scales, the mean T score for PS was 63, while the mean PK T score was 61. These results suggest that a considerable portion of the Lees-Haley subject sample appeared to exaggerate their level of emotional distress, with some subjects demonstrating a tendency to respond inconsistently to similar test items. These results indicate that patients involved in litigation are more likely to produce MMPI-2 profiles suggestive of significant emotional distress and a tendency to make a higher than average number of physical complaints, relative even to psychiatric patients and those patients with serious physical injuries. The author conceptualizes the type of patient producing such an MMPI-2 profile as “an unhappy somatizer involved in a social context which encourages rationalization, projection of blame, and complaining”, and suggests that “perhaps litigation provides a form of personal validation" for such patients (p. 754). The etiology of the problems that engender this type of MMPI-2 profile are not clear, but this research suggests a relationship between the actual event that leads the patient to pursue litigation, the stress of the litigation process, the possible payoff associated with a decision in favor of the plaintiff, and the degree to which patient complaints are produced by sampling bias. The author recommends additional research that incorporates psychometric and social-cognitive variables, in an effort to more fully illustrate the factors that operate in these cases.

1.5. Assessment Methods for the Detection of Malingering

The evaluation of suspected malingering has been accomplished by different means, with some clinicians advocating the analysis of test score patterns and inconsistencies in behavior and self-report (Williams, 2002; Martens, Donders & Millis, 2001; Slick, Iverson & Green, 2000; Slick et al., 1994; Suhr et al., 1997), or the use of specific measures of response bias (Rose, Hall & Szalda-Petree, 1995; Rose, Hall & Szalda-Petree, 1998; Hiscock & Hiscock, 1989; Binder & Pankratz, 1987), or both (Trueblood and Schmidt, 1993). The following sections discuss these approaches in detail.

1.5.1. Symptom Validity Testing

The concept that subjective reports of cognitive problems may persist in some patients, often inexplicably, after mild head injury is not of recent origin. Miller first used the term "compensation neurosis" in 1961 to suggest that some litigating patients presenting with persisting cognitive complaints were prone to conscious and/or unconscious exaggeration of their symptoms, and were not manifesting the residua of true organic impairment. This opinion was not universally shared at that time; with a subsequent survey of neurosurgeons (Auerbach, et al., 1967) suggesting that opinion was almost evenly divided between an organic versus an emotional etiology for the persistence of such symptoms. Only a small proportion of this sample of physicians (nine percent) believed that patients' persistent cognitive complaints were driven by the promise of financial compensation. At that time, systematic methods to more fully examine the incidence, and impact, of feigned cognitive impairment were not available. However, acknowledgment of this problem encouraged the development of more refined methods to detect simulated cognitive impairment, rather than relying solely on clinical judgment derived from an examination of patient performance on standard assessment instruments. Some investigators (Trueblood & Schmidt, 1993; Iverson & McCracken, 1990) have noted that the use of traditional neuropsychological tests as measures of malingering has been of inconsistent success. Inherent in the construction of such tests is the assumption that all subjects to whom the instrument is administered will attempt to perform to the best of their ability or, at the very least, will not consciously attempt to perform poorly (Slick, et al., 1994). To remedy this problem, the development of tests specific to the detection of malingering and suboptimal performance has become a major area of interest in Neuropsychology. The ability of such tests to generate a single score was seen as a favorable alternative to the use of more sophisticated analyses of score patterns on traditional neuropsychological tests, most often of memory (Rose, Hall & Szalda-Petree, 1998).

An early attempt to construct a measure of effort resulted in the 15-Item Visual Memory Test (Rey, 1964). This simple test allows the patient 10 seconds in which to view a card on which 15 items are arranged in five separate rows (capital letters A, B, C; Arabic numerals 1, 2, 3; lowercase letters a, b, c; a circle, square and triangle; and Roman numerals I, II III). Immediately after the 10-second time limit, the card is removed from the patient's view and the patient is asked to draw, on a separate sheet of paper, as many of the test items as can be recalled. As with the majority of tests of effort, the underlying principle of the 15-Item test is that those patients who desire to appear cognitively impaired will perform poorly on this test, even though, theoretically, all but the most severely impaired patients are able to recall a "minimum number" of items. This "minimum number"; however, has been the matter of much debate. Lee, Loring & Martin (1992) found that their population of neurologically compromised patients recalled seven or more items, with only seven percent of their sample recalling fewer than nine items. A subset of their sample, comprised of patients involved in litigation, recalled fewer than seven items and was also significantly less educated than the remainder of the sample. Schretlen et al. (1991) administered the 15-Item Test to a sample comprised of psychiatric and neurological patients, normals instructed to simulate such disorders, and suspected malingerers. The suspected malingerers recalled no more than eight items, fewer than any of the other subject groups in this study. Millis and Kler (1995) achieved similar results, and proposed a cutoff score of seven items. In an effort to reduce the number of false positive classifications found in other studies, Arnett, Hammeke & Schwartz (1995) used number of rows recalled, rather than single items, as a means by which to evaluate patient performance in subjects instructed to feign cognitive impairment. Two samples were used, the first being a group of college students and the second, a group of first year medical students who were felt to provide a more sophisticated example of would-be malingerers. These were compared to patients with documented intracerebral hemorrhage. Using the cutoff criteria of two correct rows in the correct location, they found a sensitivity of 47 percent and specificity of 97 percent in the college sample, and a sensitivity of 64 percent and specificity of 96 percent in the medical student sample. Two of the neurological patients were misclassified as malingerers; however, the authors proposed that visual acuity and visuospatial problems accounted for these patients' performance and cautioned against the use of the 15-Item Test in such populations.

Additional limitations are seen in those studies evaluating the effect of intelligence on Rey 15-Item performance. Goldberg and Miller (1986) found that, while their entire psychiatric patient group recalled at least nine items, approximately 38 percent of a severely mentally retarded group performed below the cutoff. However, unlike many patients who malinger or who have true memory impairment, the errors committed by the severely retarded subjects consisted mostly of perseverative responses and rotations of the test items. The results from the Schretlen et al. study (1991) demonstrated sizeable correlations between performance on the 15-Item Test and measures of intellectual ability. Such findings suggest that the 15-Item Test does not possess sufficient sensitivity and specificity for clinical effectiveness in the detection of malingering, and Millis and Kler (1995) recommend that it not be used in isolation in situations where the detection of malingering is a major thrust of the evaluation.

While most tests of symptom validity focus on the examination of feigned memory impairment, the Symptom Validity Technique (SVT) (Pankratz, Fausti & Peed, 1975) was originally developed to test the validity of a patient's complaint of sensory loss. Rather than being a test composed of specific items, presented in a preordained order, the SVT is a method of examination that allows the clinician to submit the patient's specific complaint to the rules of chance. The technique requires that the patient be presented with 100 forced-choice, two-alternative trials based on the patient's specific complaint (i.e. loss of feeling in a limb, visual problems, memory loss). After each trial, the patient is informed of the accuracy of their response. It is theorized that, in response to positive feedback, those patients who are attempting to malinger will increase the number of incorrect responses they make. In theory, such an alteration in response bias actually demonstrates functional ability within the domain being tested, as the patient must accurately perceive the stimulus in order to decide to provide an alternative response. Although the patient is told that the test is basic, but very difficult, the laws of chance dictate that approximately 50 percent of the patient's responses should be correct. Performances below chance are; therefore, suggestive of a willful attempt to appear impaired with respect to the domain being tested. While this approach is attractive, due to its simplicity and the freedom provided to the examiner to construct an appropriate task on an as-needed basis, clinicians have been cautioned about interpreting the results of such procedures, in that specificity is not assured (Larrabee, 1990).

In 1989, Hiscock and Hiscock developed a different approach to the assessment of suboptimal performance that was designed to appear more difficult than the measures used at that time. In this format, also a two-alternative, forced-choice test, the patient is briefly presented with a written number string, which appears difficult to remember. Subsequently, the patient is presented with two written number strings of the same length and is asked to identify the target string, which is easily discriminated from the foil by recognition of the first or last number. The test is presented to the patient as "very difficult", although its recognition format and the ability to choose between only two response alternatives actually make it quite simple, as the subject has a 50 percent chance of choosing the correct response each time. In addition, and unlike the symptom validity tests that predated it, the Hiscock method allows for the calculation of a chance-level score, around which confidence intervals can be calculated. These properties make it easier for the clinician to collect quantitative data, and to make more accurate inferences regarding the patient's performance. Those patients who perform at less than chance levels are generally suspected of intentionally altering their performance so as to appear more impaired, since the ability to choose the incorrect response implies that the patient's memory is sufficiently intact to allow recognition of the correct response, and to opt for the alternative, incorrect response.

Despite this refinement of the symptom validity testing procedure, the Hiscock method appeared to most normal subjects to be much easier than it was described during its administration. In an attempt to remedy this situation, and to add to the face validity of the measure, Slick et al. (1994) modified the Hiscock test by manipulating the apparent difficulty of some of the test items, identifying this version as the Victoria revision of the original Hiscock procedure. In addition, the number of test items was decreased, and the task was administered on a computer rather than on cards. Their results demonstrated a relatively low correct classification rate for subjects instructed to malinger; however, when only the difficult test items were considered, 83 percent of subjects were correctly classified. In addition, a more educated group of simulated malingerers was classified correctly. The authors reported that the results from their TBI group indicated that the Victoria revision was sensitive to real memory impairment; however, their TBI sample was also litigating, thus confounding these results.

Further refinement of the forced-choice method resulted it the development of the Portland Digit Recognition Test (PDRT; Binder and Willis, 1991) which, unlike its predecessor, incorporates verbal presentation of stimuli followed by a delay interval of a pre-prescribed length, during which the patient is required to count backwards from a specified number. After this delay interval, the patient is required to discriminate the target number string from a foil. The introduction of the distraction factor in this test is believed to increase its specificity as a measure of memory functioning, as amnestic patients generally cannot recall previously presented information after a delay or distraction, even when the delay or distraction is quite brief (Baker et al., 1993). The Baker et al. study also demonstrated that college students who had been instructed to simulate memory impairment generated scores below normal limits irrespective of the use of the distraction trial. In contrast, amnestic subjects generally performed at the same level as normals when the distraction task was removed from the test paradigm. These differences in patterns of performance were seen as valuable in aiding in the diagnosis of true memory impairment as well as in the detection of deliberately poor performance. However, the considerable length of time required to administer the task is inefficient, at best, and may negatively impact a patient's attention and frustration tolerance, especially if the task proves to be quite easy (Lezak, 1995). Patient performance on the PDRT has also been found to be susceptible to the effects of abnormal attention and impaired immediate recall, thus reducing its sensitivity as a measure of malingering (Binder & Willis, 1993). In order to offset the possibility that more sophisticated subjects would be able to accurately scrutinize, and thus modify accordingly, their own performance on overt measures of malingering, Rose et al. (1995) developed a computerized version of the Portland Digit Recognition Test (hereafter referred to as the PDRT-C) that also recorded the subject's response latency. Their results demonstrated that adding measurement of response time greatly improved their ability to correctly identify simulated malingerers.

An additional test of malingering, the Test of Memory Malingering (TOMM; Tombaugh, 1996) was based on the use of a recognition paradigm, despite the problems noted previously with the use of such a relatively simple test format. The TOMM consists of 50 line drawings of common objects that are presented individually to the patient for three seconds each. In cases where the patient is suspected of being inattentive, the test manual (pg. 17) instructs the examiner to point to each picture as it is presented. The patient is instructed to look at each picture and attempt to remember it. After presentation of the pictures, the patient is presented with the 50 target pictures; each one paired with a foil, and is asked to identify the target. For each selection, the patient is informed whether their choice is correct or incorrect and, for incorrect responses, the examiner points to the correct picture. This procedure is repeated on a second learning trial, which is administered immediately after Trial 1 and in which the target pictures are paired with a different set of foils than those found in Trial 1. An optional Retention Trial can be administered 15 minutes after the completion of Trial 2, in which the patient is presented with the 50 target pictures, paired with yet a different set of foils, and is instructed to select the appropriate target.

Validity studies (Tombaugh, 1996) suggest that the TOMM is robust to the effects of various forms of severe cognitive impairment. Of all the clinical groups, those patients diagnosed with dementia generated the lowest scores, with the lowest score being 33 correct responses on Trial 2 in two patients judged to be moderately to severely demented, based on their performance on the Mini-Mental Status Examination. A sample of non-litigating brain injured patients generated mean scores of 45, 49 and 50 on each of the TOMM trials, respectively. However, a sample of 11 litigating brain injured patients generated mean scores of 25 correct on Trial 1, 33 correct on Trial 2 and 35 correct on Retention, equal to or below those scores of patients with advanced dementia. Therefore, these studies appeared to demonstrate the utility of the TOMM as a useful measure of malingering.

The Word Memory Test (hereafter referred to as the WMT; Green, Allen & Astner, 1996) was designed as an ostensible measure of verbal learning and memory, but was specifically devised as a measure of response bias, according to the recommendations of Nies & Sweet (1994). These guidelines suggest that malingering is associated with relatively low scores on tests of verbal learning and recall, recognition memory scores that are worse than scores on free recall measures, inconsistent test performances, and scores on symptom validity tests that are below established cutoff points or are below chance levels. In constructing the WMT, the authors sought to create an instrument on which even the most severely brain injured patients would score above 90 percent correct. Accordingly, a score that is three standard deviations below the mean for such a group is still above a chance level of responding, meaning that scores that fall below this three standard deviation cutoff point would be highly unusual and would strongly suggest biased responding (Green, Allen & Astner, 1996). The WMT consists of 20 semantically related word pairs, 10 of which are considered "easy" (i.e. "dog-cat") and 10 that are considered more difficult (i.e. "tree-lake"). This list of word pairs is read to the patient in two consecutive learning trials. Subsequently, the patient is presented with an Immediate Recognition Trial, during which each target word is paired with a semantically related foil (i.e. "dog-rabbit", "cat-mouse", "leaf-tree", "stream-lake") and is asked to denote the target word. Unlike the TOMM, the patient is not informed of their errors. A Delayed Recognition trial, presented in the same format as Immediate Recognition, but with different foils, is administered 30 minutes after the completion of Immediate Recognition. A Consistency Index is also calculated, which tabulates the degree to which the patient provided the same correct response on Immediate and Delayed Recognition. A computerized version of the WMT is also available. Research using the WMT has supported its ability to discriminate between litigating patients with mild versus severe brain injuries, with mildly brain injured litigants performing below their severely injured counterparts, and often below chance levels (Green, Iverson & Allen, 1999).

The Computerized Assessment of Response Bias (CARB; Allen et al., 1997) was designed as a quantitative method of examining response bias, and uses a forced-choice digit recognition format. As its name implies, it is computer administered in an effort to reduce administration and scoring time on the part of the examiner. The test is presented to the patient as a "visual digit memory test", and consists of three "blocks" of stimuli.

The patient is also informed that progression to subsequent blocks is contingent on adequate performance, although this is not the case. The supposition in this format is that those patients who are attempting to malinger will assume that their performances are adequate, and will intentionally perform even more abnormally on subsequent blocks, resulting in highly unusual scores.

The relatively large number of symptom validity tests that have been developed in the past several years underscores the emphasis that is being placed on this facet of neuropsychological assessment. However, the tests that have resulted from this development have also been criticized for a number of shortcomings that are felt to undermine their utility as measures of malingering. Inherent in the instructions for many of these tests are suggestions that the tasks are quite difficult when, in fact, the tasks are designed to be simple enough for even the most cognitively compromised patients to perform well. Their relatively simple nature is believed to reduce their face validity, and may not convince the more sophisticated or informed subject that the test is simply a measure of memory (Slick, et al., 1994). In addition, some tests, such as the Portland Digit Recognition Test (PDRT) are quite lengthy, and the addition of several such instruments can considerably extend the administration time of the entire test battery (Tombaugh, 1996).

Critics of the forced-choice method of malingering assessment suggest that many of these tests are insufficiently sensitive, and therefore, result in a high number of false negative classifications. Many symptom validity tests require that the patient generate a worse-than-chance performance in order to be classified as malingering, relying on the binomial theorem for the derivation of cutoff scores. This relatively simple test format does not require that normative samples be collected for these tests, as simple cutoff scores could be calculated based on the number of responses that constituted a chance level of effort (Williams, 1998). However, it has been demonstrated that the simplistic nature of these tests resulted in lowered sensitivity (Binder, 1993). Iverson and Binder (2000, p. 841) caution, “Any FCT (forced-choice test) that relies on binomial significance and that lacks empirical derivation of a cutoff score is inadequately sensitive for routine clinical use”. In order to correct for the potential classification problems raised by the use of binomially derived cutoff scores, some of the newer symptom validity measures have cutoff scores that are empirically derived, in an effort to increase sensitivity and enhance their clinical utility. In addition, the methods used to validate many of the symptom validity tests have been questioned. As with many other psychological and neuropsychological measures, validation of symptom validity measures on the specific population targeted for assessment is highly recommended (Coleman, et al., 1998). Despite these advances, such problems continue to raise questions about the utility of many of the malingering measures currently in use.

1.5.2. Unusual Response Styles

Tombaugh (1996, p. 4) notes “consensus exists among neuropsychologists, forensic psychologists, and psychiatrists that inconsistency is the hallmark of malingering”. Iverson and Binder (2000); therefore, suggest that the clinician be cognizant of any inconsistencies that occur during the evaluation, both in test performance as well as in the patient’s behavior. Often, these take the form of a patient who is able to provide a detailed history of their accident or illness, subsequent treatments and course of recovery, but whose memory test performance is significantly impaired. Other inconsistencies cited by the authors include patients who complain of problems with receptive/expressive language and perform poorly on formal tests of such functions, but who are able to carry on conversation with no apparent limitations. Inconsistencies in motor performances may also be seen, such as patients who complain of loss of upper extremity function but are able to write, draw and gesture without apparent problems. Patients may also perform inconsistently on different tests of the same cognitive construct. Finally, the authors suggest that clinicians examine inconsistencies across serial evaluations, with special attention to scores on the same tests that decline from one evaluation to the next. Such findings would be inconsistent with the expected recovery from brain injury over time, as well as the effects of practice.

Greiffenstein et al. (1994, 1995) provide guidelines for the empirical examination of performance inconsistencies, and suggest that subject samples be divided not on the basis of their litigation status, but on the basis of other criteria that are typically part of the clinical evaluation. In particular, the incidence of such factors such as poor performance on at least two neuropsychological tests, total disability in one of the patient’s major social roles, lack of agreement between the patient’s reported problems and collateral information, and loss of long-term memory often suggest an attempt on the patient’s part to feign cognitive impairment and a level of disability deserving of financial compensation. In their samples, patients who met at least two of these criteria generated more abnormal scores on memory tests than those patients with persistent memory complaints who did not have any of the problems noted above. Interestingly, the suspected malingering group’s overall performance on memory tests was comparable to that obtained by patients with more severe brain injuries, thereby providing a more refined and specific classification than is provided in other studies using a similar format.

Patients who consciously attempt to produce impaired performances on neuropsychological tests may use different methods to achieve their goal. According to Williams (1998), a patient's performance on features that are common to many standard memory tests, such as immediate recall (i.e. Digit Span) and recognition trials, can be analyzed to detect patterns of performance that are illogical and; therefore, suggestive of malingering. Accordingly, malingerers may employ different response styles in an effort to appear impaired with respect to memory. For example, patients may choose to respond incorrectly, based on their ability to estimate the number of incorrect responses that can be given without creating the impression that their poor performance is intentional. The ability to do so depends on the complexity of the test in question, with those tests that are composed of a specific number of trials that can be counted or anticipated being the easiest on which to malinger. Patients may also feign problems with attention and concentration, or may respond very slowly, thus affecting their scores on timed tasks and producing the illusion of cognitive dysfunction. As noted previously; however, response latency has been shown to discriminate the performances of malingerers and controls on the Portland Digit Recognition Test (Rose, et al., 1995); therefore, observation of such behavior may suggest that the patient is performing below their capability. Malingering patients may also respond in an indiscriminant manner, especially on tasks that allow the subject to choose an answer from a set of possible responses (i.e. recognition tests). Finally, some patients may respond in an incorrect, but systematic fashion, such as providing an intentionally incorrect response on every third trial. Consideration of such response styles is; therefore, recommended when considering the validity of the patient's performance.

1.5.3. Pattern Analysis

Suhr and Gunstad (2000) advocate the use of an analysis of score patterns on traditional neuropsychological measures for the detection of malingering, based on several points. They point to the often-transparent nature of many malingering measures, which allows the relatively sophisticated malingerer to adjust their performance accordingly to avoid detection. In addition, they note that the addition of multiple symptom validity measures can add considerable administration time to an already lengthy test battery. For these reasons, they suggest that more traditional tests of cognitive functioning be used as measures of malingering; however, they caution against the use of cutoff scores that are often derived from these measures, citing previous findings that such cutoff scores do not reliably discriminate malingerers from those patients with true organic impairment.

Iverson and Binder (2000) also advocate the use of pattern analysis, and describe a systematic method for analyzing such patterns on standard neuropsychological tests, using information gathered from the considerable body of research that has been conducted with regard to the establishment of cutoff scores for acceptable performances on such tests. The authors' approach involves four steps, each considering the base rates for specific test performances in different populations, including normal subjects, non-litigating brain injured patients, patients with severe brain injuries, and known or suspected malingerers. Larrabee (2000) reminds clinicians of the need to remember that symptom validity must be established for all patients seen in the course of forensic evaluation, as it cannot be assumed that patients who have actually sustained injuries will not attempt to exploit them for the promise of increased external rewards, particularly money. He goes on to note that, for many mild brain injury patients involved in litigation, patterns of performance on neuropsychological tests are often illogical, and do not fit with those patterns that have been established in the literature.

Patterns of performance on list learning tasks such as the Rey Auditory Verbal Learning Test (RAVLT) and the California Verbal Learning Test (CVLT) have also been examined for evidence of malingering. Results have generally demonstrated that patients suspected of malingering are more likely to generate scores on the recognition trial that are significantly below those generated on the free recall trials (Bernard, 1996; Millis et al., 1995; Bollich et al., 2002). In addition, malingering subjects were found to recall fewer words from the first third of the RAVLT list than controls, generating a serial position effect that was believed to discriminate malingering from non-malingering subjects (Bernard, 1996). In another study using the RAVLT, Suhr et al. (1997) found that their group of suspected malingerers performed similarly to other groups with respect to learning on trials one through five; however, their total number of words acquired was significantly below that of head injured, litigating patients, depressed patients, and patients diagnosed with somatoform disorders. In addition, the suspected malingerers performed below these same groups on RAVLT delayed recall trial and on numbers of recognition true positive and true negative responses. Finally, the suspected malingerers were more likely than these same groups to fail to recognize a word that had previously been recalled at least three times on the five learning trials.

In an effort to investigate the utility of previously established cutoff scores designating valid performance on the CVLT, Slick, Iverson and Green (2000) examined 193 patients seeking compensation for alleged brain injuries. By using cutoff scores for four of the CVLT indexes (Total Trials 1-5, Long Delay Cued Recall, Recognition Hits and Recognition Discriminability), they constructed a Summary Index which was assigned a positive value if a suboptimal score was found on any one of the four subtests from which it was comprised. The Summary Indexes were compared between the two groups. The results indicated that outcome was unrelated to the age of the patient and the severity of the brain injury. However, in contrast to their hypotheses, moderately to severely impaired patients in their sample performed worse than less severely impaired patients on the CVLT, resulting in a higher than average false positive rate for the classification of suboptimal performance using this composite index. The authors suggest that the CVLT is sensitive to true memory impairment, consistent with its intended use as a neuropsychological measure, and recommend the use of more conservative cutoff criteria (Recognition Hits and Discriminability) when using this test in the assessment of malingering. Connor, et al. (1997) examined the utility of the forced-choice recognition task on the CVLT. Using a cutoff of less than 87 percent correct, they were able to correctly classify simulated malingerers with 80 percent sensitivity and 97 percent specificity, and suggested that this test was as useful in the detection of insufficient effort as other forced-choice format tests, but could be administered in less time, making it a more efficient assessment method. With respect to the CVLT-II, the test manual notes that more than 90 percent of the normative sample achieved a perfect performance on the forced-choice recognition trial, theoretically making this test ideal for the detection of insufficient effort. However, research on the true utility of this test has yet to be reported.

Previous research using the Wechsler Memory Scale-Revised (hereafter referred to as the WMS-R) in the assessment of malingering has been relatively sparse. Mittenberg et al., (1993) demonstrated a difference between the General Memory and Attention/Concentration Indexes that correctly classified 83 percent of their sample of simulated malingerers. In 1996, Iverson and Franzen employed a two-alternative forced-choice recognition trial for the Logical Memory I subtest of the Wechsler Memory Scale – Revised. Using a cutoff score of less than 18 (out of 24) correct forced choice responses, a correct classification rate of 93 percent was achieved for their group of experimental malingerers, and a 97 percent overall correct classification rate when all groups were considered.

Even fewer studies exist that examine the utility of the WMS-III in the detection of malingering. This is surprising, given that some of the subtests of the WMS-III (i.e. Logical Memory, Word Lists, Verbal Paired Associates, Visual Reproduction and Faces) have been expanded to include two-alternative, forced-choice recognition trials, ostensibly making them ideal candidates for the examination of malingering on this test. Killgore and DellaPietra (2000a) examined the wording and ordering of questions on the recognition trial of the Logical Memory subtest, hypothesizing that the wording of some of the questions would bias respondents toward a “yes” or “no” response set, while other questions primed respondents toward a “yes” response on the next question in the subtest. These hypotheses were confirmed in their study, in which the Logical Memory recognition subtest was administered to a group of 31 normal subjects and 36 neurologically impaired patients who had never heard the two Logical Memory stories. The results identified a subset of six items on the recognition subtest that were rarely answered incorrectly, even by naïve subjects. This subset of items was termed the Rarely Missed Index (hereafter referred to as the RMI). In a follow-up study, Killgore and DellaPietra (2000b) examined the same effect in a group of 36 simulated malingerers and in data collected retrospectively from a group of 51 non-litigating patients with a variety of neurological problems. Their results indicated that the RMI achieved a sensitivity of 97 percent and a specificity of 100 percent in the sample used in this study, comparable to the efficacy of many of the more recognized malingering tests currently available. The RMI was also endorsed for its efficiency as a test of malingering, since the Logical Memory recognition subtest is an inherent part of the WMS-III battery and; therefore, the ability to calculate the RMI is not dependent on the administration of an additional test.

1.5.4: Comprehensive Methods

While the above discussion presents evidence to support individual methods of malingering detection, by the use of specific tests, analysis of behavioral styles, or recognition of specific patterns of performance on cognitive tests, no one method has emerged as the “gold standard” of malingering detection. Nies and Sweet (1994) proposed a multidimensional approach to malingering detection which combined the results of specific malingering tests, results of standard neuropsychological measures administered in forced-choice formats, consideration of both intra- and inter-subtest scatter and illogical patterns of test performance, and inconsistencies between test performance, patient self-report and collateral data. While this method provided suggestions for a better-rounded conceptualization of cases in which malingering was suspected, it lacked a systematic method by which to apply these concepts. Slick, Sherman and Iverson (1999) refined this approach, providing a definition of malingering of cognitive dysfunction in the context of forensic neuropsychological assessment, and organizing a set of diagnostic criteria into the form of a comprehensive and logical decision-making strategy for the detection of malingering of neurocognitive deficits in such cases. Central to these diagnostic criteria is the concept that diagnostic certainty in the detection of malingering is not possible; therefore, these criteria provide methods of information analysis that provide levels of diagnostic certainty, rather than forcing the clinician to classify a given patient in a dichotomous, “all-or-nothing” manner. Accordingly, classifications of the likelihood of Malingering Neurocognitive Dysfunction (hereafter abbreviated “MND”) can be made, including designations of Definite, Probable and Possible MND. Criteria to be considered include 1) the presence of a substantial external incentive, which is required for any level of diagnosis to be made; 2) evidence from neuropsychological testing, including, but not limited to, discrepancies between known patterns of brain functioning, various patient behaviors, information from collateral reports, and the test data; 3) evidence from the patient’s self-report, including discrepancies with the documented history, behavior, collateral information and evidence of exaggerated or feigned psychological problems; and 4) that the behaviors meeting the necessary criteria from points 1 and 3 are not better accounted for by other (psychiatric, neurological or developmental) factors. While this method of decision-making and diagnosis was intended to provide a framework to inform the clinician’s conceptualization of an individual patient, the authors warn that the proposed criteria cannot be applied rigidly to every case, and that clinical judgment is also necessary.

1.6. The Effects of Coaching on Neuropsychological Test Performance

When patients present for neuropsychological assessment, it is reasonable to expect that, while they may have been previously exposed to some or all of the tests used in such an assessment, they are unaware of the intent and/or the structure of these specific tests. However, within the area of forensic assessment in Neuropsychology, there is growing evidence that some patients arrive for testing with intimate knowledge of the test instruments, ostensibly provided by their legal counsel. Of course, such practices pose serious threats to the integrity of such evaluations, certainly compromise the validity of the test results, and expose proprietary and copyright-protected information to the general public.

The absolute degree to which attorneys provide information to their clients is unknown; however, prior survey research (Wetter & Corrigan, 1995) indicated that 65 percent of both law students and attorneys reported that they believed that they should inform their clients of the purposes of psychological testing. In addition, 36 percent of the law students and 48 percent of attorneys in this study stated that they felt that they should always or usually inform their clients that tests of effort and symptom validity may be administered in the course of a neuropsychological evaluation. These data are concerning, particularly since they were collected almost 10 years ago and may not accurately reflect the increase in sophistication that has presumably occurred with respect to attorneys' knowledge of such issues, and their attempts to coach their clients accordingly. It has also been demonstrated that the provision of education and information regarding the cognitive sequelae of brain injury allows subjects instructed to feign cognitive deficits to produce test scores that are similar to patients with acquired injuries (Frederick et al., 1994).

More recently, Essig et al. (2001) conducted survey research to examine the current practices of forensic neuropsychologists and attorneys. Their results indicated that approximately 75 percent of the attorneys sampled actively prepare their clients prior to the assessment. In these cases, "preparation" often included dissemination of specific test content. Approximately eight percent of the attorneys responding to the survey admitted that they instruct their clients on how to respond to specific questions on cognitive measures. In addition, the results indicated that a large proportion of the responding attorneys reviewed specific malingering tests with their clients, as well as the validity indicators on more traditional tests. For example, approximately 29 percent of the attorneys surveyed reviewed the MMPI-2 with their clients. In addition, the authors note that, even if the attorney does not review a specific malingering test with the client, the information provided may allow the client to recognize typical symptom validity test formats and to adjust their test performance accordingly. These findings suggest that the practice of client coaching by attorneys should be recognized as a growing concern in the field of forensic neuropsychological assessment.

In addition to the continued threat of attorney coaching, concerns are also being raised regarding the availability of information about neuropsychological testing in the popular media, and on the Internet. A recent article in the Boston Globe (Barry, 2002) reviewed a recently published article pertaining to the use of diagnostic and symptom validity testing, and the availability of such information on the Internet. Ironically, the publication of such an article in the popular press would appear to only exacerbate problems with test security, as interested newspaper readers would be alerted to the ready availability of such information. The study in question, by Ruiz, et al. (2002), undertook an extensive search of Internet websites; using standard search engines, and revealed a multitude of sites devoted to the dissemination of information relevant to neuropsychological assessment. The majority of these sites (70 to 85 percent) were considered to be of only minimal threat in terms of their ability to provide information that would allow someone to successfully dissimulate cognitive impairment. Approximately 20 to 25 percent of these sites were judged to be indirect threats to test security, as they presented enough information to indirectly threaten the validity of certain psychological tests. Two to five percent of the sites; however, were considered direct threats in that they presented detailed information about specific tests, including test questions and examples of test stimuli (i.e. the stimuli from the Mattis Dementia Rating Scale as well as the Rorschach inkblots). One "popular Internet auction site" (Ruiz, et al., 2002, p. 296) listed a set of Rorschach cards for sale. Still other sites were found to present detailed information regarding the methods used to detect malingering and dissimulation on several tests, including a discussion of the validity scales of the MMPI-2. Still other sites were found that provided detailed information regarding the purpose of medicolegal evaluations, with suggestions on how to conduct oneself in a convincing manner so as to obtain disability benefits. These results are disturbing, at best, particularly in that the authors noted that the majority of the developers of the sites deemed to be direct threats identified themselves as lawyers or, even more concerning, as psychologists. Even before this study was published, such threats to the integrity and usefulness of tests of malingering had been reported. In 1989, Coleman et al. suggested that, given the increasing sophistication of legal professionals who are concerned with their clients' performance on neuropsychological examinations, and who coach their clients accordingly, the usefulness of many of the symptom validity measures in current use may be time limited and that more, and more sophisticated measures, are in need of development.

In order to more fully explore this topic, a multitude of studies have been undertaken to examine the effects of coaching on subject performance on both tests of malingering and on traditional measures of cognitive functioning. Since a sample of confirmed malingerers is generally difficult to come by, many studies have utilized subjects instructed to simulate the perceived effects of acquired brain injury. However, Suhr et al. (1997) note that studies that rely on simulation of malingering may not be generalizable to those situations that the practicing neuropsychologist encounters in the normal course of their clinical work, highlighting their limited utility as a source of information regarding the detection of malingering. Despite these limitations, a review of the pertinent literature on the effects of coaching is presented below.

Coleman and colleagues (1998) examined the effect of coaching on performance on the CVLT. Their sample of 90 undergraduate students was divided into three groups, consisting of subjects instructed to malinger, subjects instructed to malinger and also provided with information regarding the cognitive sequelae of mild head injury, and controls. While their results indicated that the use of some CVLT indices, such as Total Learning and Delayed Recall, is useful in detecting patients who may be malingering, those subjects who are coached are often able to modify their responses such that their scores on these measures are within normal limits. However, their work suggested that more subtle indices, such as Discriminability, Recognition Hits and learning slope, are relatively robust to the effects of coaching. Their findings are consistent with, although somewhat less definitive than, the findings in previous work by Millis et al. (1995); which indicated that Recognition Hits provided the most predictive power in their attempts to correctly classify naïve and coached malingerers. They recommend that special attention be given to a patient's Recognition Hits score in cases where malingering is suspected.

As noted previously, Suhr and Gunstad (2000) have recommended that clinicians forego the use of specific tests of malingering in favor of the analysis of score patterns on traditional neuropsychological tests due, in part, to the obvious nature of many of the malingering tests. They examined the performances of naïve, coached, and subjects who were coached but also warned about the likelihood that their deceptive test taking strategies would be discovered, on the RAVLT. The authors hypothesized that coached malingerers would perform poorly on all tests, that coached but warned malingerers would score within normal limits on forced-choice malingering measures (i.e. the PDRT), but would generate impaired scores on the RAVLT, and normal controls would perform well on all tests. Their results demonstrated that, despite warnings, the coached and warned subjects generated suspicious patterns of performance on Learning versus Recognition scores on the RAVLT that revealed their intent to malinger. However, this group was able to alter their collective performance on the PDRT such that the sensitivity of this measure as a test of malingering was reduced. The authors concluded that reliance on a pattern analysis method appears to be robust to the effects of both coaching and warnings to subjects about the possible use of methods to detect feigned memory impairment.

Sullivan, Keane & Deffenti (2000) offered incentives to two groups of subjects instructed to malinger memory deficits on the Rey Auditory Verbal Learning Test (RAVLT), with one group warned that procedures to detect insufficient performance would be used. They hypothesized that the group receiving the warning would perform better than the no-warning group; however, their analysis revealed no differences in overall performance between these two groups, but also revealed that both groups performed below a group of controls. Follow-up questioning of the warned subjects revealed that, despite the warning, these subjects felt confident that they could generate false, yet believable, memory deficits on the RAVLT. The authors concluded that the coaching strategies used in this study, which included the promise of an incentive for believable but faked memory deficits, were highly effective despite the possibility that this false performance might be detected.

The effect of coaching on the Wechsler Memory Scale-Revised has also been examined empirically (Johnson & Lesniak-Karpiak, 1993). The results of this study indicated that coached subjects scored below normal controls on the Visual Memory, General Memory and Delayed Memory indices, whereas those subjects who were coached, but warned about the use of detection methods, scored similarly to normal controls on these measures. The authors subsequently replicated their findings (Johnson & Lesniak-Karpiak, 1997). Bernard (1990) used an incentive strategy to examine differences between compensated and uncompensated malingerers and controls on the

WMS-R, the Rey Complex Figure, and the RAVLT. The data demonstrated no significant differences between the malingering groups; however, both of these groups' scores on these tests were significantly below those of the control group.

Other studies have examined the effect of coaching and warning on subjects' performance on specific tests of malingering. Rose et al. (1998) found that the computerized Portland Digit Recognition Test identified approximately 70 percent of simulating malingerers; however, Suhr and Gunstad (2000) found that warning subjects about this detection strategy reduced the sensitivity of this test to approximately 53 percent. Rose et al. (1998) also found that coached malingerers performed significantly better than uncoached malingerers on other tests of response bias, namely the Nonverbal Forced Choice Test (Frederick & Foster, 1991) and the 21-Item Test (Iverson et al., 1991). A similar study using a revised version of the Hiscock and Hiscock forced-choice memory task found that this test was not sufficiently sensitive to detect the performance of coached malingerers who had been warned about the nature of the test (Slick, et al., 1994).

These studies raise issues surrounding the use of warnings in the context of forensic assessment. Conflicting opinions have been expressed regarding warning subjects that their attempts to malinger could be detected, with some expressing the belief that providing such warnings, and discussing the possible consequences, provides concrete information regarding the intent of the symptom validity portion of the evaluation. Those in favor of such a procedure state that such disclosure practices are consistent with ethical principles governing informed consent (Johnson & Lesniak-

Kerpiak, 1997). However, others feel that such warnings only serve to produce more sophisticated malingerers (Youngjohn et al., 1999).

The issues discussed above represent some of the more recent issues in the area of malingering detection. Concerns appear to be growing regarding the degree to which patients are coached by their attorneys or are more cognizant of malingering detection methods and, recently, it has been proposed that tests used to detect malingering may become less effective as individuals become more aware of their use and content. Almost no research has been done in an effort to examine such concerns; however, recently, Greiffenstein and Baker (2002) examined trends in the ability of the Rey 15 Item Test to detect malingering, hypothesizing that improvements in subject performance over time would suggest increasing sophistication of subjects taking this test. While not finding absolute evidence of pervasive coaching on this test, their results did suggest trends in improved performance over time in patients with late onset postconcussive symptoms. These results were interpreted as suggestive of an increase in the number of subjects who were familiar with the intent of the test and its content. Considering these results, it is reasonable to suggest that similar research using other traditional malingering tests is needed.

1.7. Special Considerations in Malingering Research

According to Cercy, Schretlen and Brandt (1997, p. 89), "no 'gold standard' exists to determine with certainty whether a patient is feigning symptoms independent of the clinical tests or experimental procedures being investigated". Attempts to investigate malingering have been hampered by methodological problems, and the various methods commonly used in malingering research continue to be the topic of significant and ongoing debate. For example, simulation research has been employed frequently in the investigation of malingering. These research paradigms are closely related to traditional experimental methods, wherein subjects are assigned to experimental and control groups. Generally, the experimental group receives some instruction to malinger, and may also receive information on symptoms associated with acquired brain injury. Control groups have generally received only basic task instructions and encouragement to perform optimally. According to Rogers (1997), the results of these studies are often not generalizable, as the groups that are typically included in such research are not culled from the clinical settings of interest. In addition, concerns are often raised regarding the composition of the groups used in such studies. Typically, such groups are comprised of college students, given their availability and frequent desire to earn extra course credit. However, Haines and Norris (2001) suggest that the advanced academic status of the average college student subject renders them considerably different from the types of patients under consideration in malingering research, further reducing the generalizability of the obtained data.

In contrast, the use of the known-groups design in malingering research is relatively less common, as it calls for the use of actual patients with diagnosed disorders, as well as malingerers. While clinical subjects are generally accessible, the development of a criterion group of malingerers is hampered by the obvious difficulty of correct identification of such patients. Therefore, while the data collected in a known-groups design are generally more clinically relevant and generalizable, due to inclusion of known patient groups, a complete discussion of the results generated by the malingering group is limited to only those malingering subjects that were successfully detected (Rogers, 1997).

The above discussion highlights some of the current problems that are frequently encountered in malingering research. Rather than continue to advocate one method of investigation over another, the use of combined methods in a single research protocol has been suggested, in an effort to control for the limitations that are inherent in each individual design (Rogers, 1997). Future research using such combined methods may add substantively to the body of knowledge that has developed relative to the assessment of malingering and deception in neuropsychological assessment.

1.8. Aims of the Current Study

The preceding review provides an introduction to the nature of memory, its assessment, and its role in the problem of malingering detection. It should be apparent from this review that questions regarding the veracity of the patient's cognitive complaints, and the validity of their performances on neuropsychological tests, are compounded when litigation is introduced as an intervening variable. Specific aspects of this complex problem were examined in this study.

As noted previously, it has been suggested that information about specific malingering tests and other methods used to detect feigned memory impairment is becoming more readily available to the public via the media, specifically, the Internet. In addition, there have been suggestions that an unknown number of attorneys may provide their clients with specific information regarding the nature of many neuropsychological and psychological tests, including specific test content, and "coach" their clients regarding methods to avoid detection of their feigned impairments. If accurate, one would suspect that malingering detection methods might become less sensitive over time as such information becomes more widely distributed among attorneys and their clients. With this in mind, an analyses of symptom validity test scores collected over consecutive years was undertaken; with an emphasis on the detection of improvement in such test scores over time that may be due to knowledge of both test stimuli and purpose. As noted previously, to date only one study has examined changes in one symptom validity test (the Rey 15 Item Test) over time (Greiffenstein & Baker, 2002), with trends in improved performance that were believed to suggest increased sophistication on the part of the subjects. Given the retrospective nature of the data available in the Tulane University database, it was possible to determine if similar trends existed in the accumulated data from this and other symptom validity tests. Specifically, it was hypothesized that if litigating patients were indeed becoming more educated regarding specific symptom validity tests, scores on these tests should demonstrate decreased variability and a lower frequency of abnormal scores, resulting in fewer classifications of invalid performance, over time. With respect to the present data, it was hypothesized that such an effect would be seen in the data pertaining to the Rey 15 Item Test and the TOMM over the 3 years of accumulated data in the Tulane database. Such a trend was not expected in the WMT or the Delayed Yes/No Recognition and Forced Choice subtests of the CVLT-2, owing to their relatively more recent introduction as measures of malingering.

As noted previously, memory impairment appears to be the most frequently malingered, or at least, the most often exaggerated, symptom of cognitive dysfunction in cases where acquired cognitive problems are alleged. Conflicting opinions, and rationales, abound for the use of analysis of test score patterns on traditional neuropsychological tests, versus reliance on more specific tests of malingering as a means of examining patient performance, and a considerable body of research is devoted to this subject. However, definitive statements regarding the utility of already-established tests of memory are lacking, and no studies to date have examined the utility of newer memory tests (i.e. WMS-III, CVLT-II) in the discrimination of suspected malingerers from those believed to be exerting optimal effort. Such questions are of significance, as all of these tests are now widely used in standard neuropsychological practice. Multiple hypotheses regarding the utility of newer test instruments can be examined empirically, given the wealth of data generated by the inclusion of new subtests and the expansion of existing subtests in these latest versions. The literature reviewed previously has suggested that recognition memory procedures, in which subjects are instructed to discriminate between target and non-target information rather than recall the target information spontaneously, are effective methods by which to discriminate malingering patients from those attempting to perform optimally, as they are relatively easy to perform and are generally robust to the effects of even significant neurological impairment. While such patterns of performance have been demonstrated on some traditional neuropsychological memory tests, such as the Rey Auditory Verbal Learning Test and the original version of the California Verbal Learning Test, similar research with newer tests using a recognition memory format has not yet been done. Given these considerations, for the purposes of the current study it was hypothesized that litigating patients would differ significantly from non-litigating patients with respect to their scores on measures of recognition memory for verbal materials, according to their scores on the Delayed Recognition subtest of the

Delayed Yes/No Recognition Subtest of the California Verbal Learning Test-II, and the Auditory Delayed Index of the Wechsler Memory Scale – Third Edition. It was also expected that litigating patients would differ significantly from their non-litigating counterparts on tests of recognition memory for visual materials, according to their scores on the Delayed Recognition portion of the Wechsler Memory Scale – Third Edition Visual Reproduction subtest.

In addition to the hypothesized differences between litigating and non-litigating subjects with regard to recognition memory test performance, it was hypothesized that these groups would also differ with respect to their performances on a variety of commonly used symptom validity tests, which are often administered in a recognition memory format. Specifically, it was predicted that the litigating patients would obtain scores below established cutoff points on the TOMM, WMT and Rey 15 Items; therefore classifying them as malingering, whereas non-litigating patients would score above these cutoff points and would be classified as non-malingering.

The foregoing literature review indicates that patients presenting for neuropsychological assessment in the context of litigation frequently characterize themselves as cognitively impaired, with reduced memory being the most frequent complaint. It was therefore hypothesized that litigating patients would make subjective complaints of memory problems with significantly greater frequency than non-litigating patients, according to their responses to such questions on the Neuropsychological Symptom Checklist.

Prior research has also suggested that litigating patients tend to exaggerate symptoms of emotional and physical distress. Specifically, litigating patients have been found to generate MMPI and MMPI-2 profiles with significant elevations, within the Critical Range, on the F scale as well as scales 1 (Hypochondriasis) and 3 (Hysteria) with greater frequency than any other scales on this test. It was hypothesized that the same effect would be seen in the current study, with litigating patients generating profiles as described here with significantly greater frequency than non-litigating subjects. In addition, it was hypothesized that litigating patients would generate scores on the BDI-II and BAI that were significantly higher than their non-litigating counterparts, indicating that these subjects perceived themselves as more significantly depressed and anxious.

1.9. Method

1.9.1. Subjects

The subject pool was collected retrospectively from the database at Tulane University Medical Center, Department of Psychiatry and Neurology, Division of Neuropsychology. The available data represented a subset of the pool of sequential adult cases referred for outpatient neuropsychological evaluation in the Tulane University Neuropsychology Laboratory, with referrals encompassing personal injury cases (from defense and plaintiff attorneys as well as Workman's Compensation), non-forensic medical evaluations, psychoeducational evaluations, and other evaluations that do not fall into these categories. The available database (using Microsoft Access for Windows '97) consists of test data from approximately 500 subjects seen for neuropsychological assessment between 1998 and 2002. These data have been coded according to a system developed collaboratively by F. William Black, Ph.D., this author, and the staff psychometrist in the Division of Neuropsychology. In order to ensure patient confidentiality, each patient's data were identified by code number only, consisting of, in order, the date of evaluation (two digit year, month, and date) and an additional number identifying the examiner (1 = postdoctoral fellow, 2 = intern, 3 = psychometrist).

1.9.2. Measures

The specific test instruments included in the TUMC Neuropsychological Battery are commercially available measures that are in common use throughout the country. Therefore, the available data encompass multiple cognitive, neurobehavioral and emotional domains, including intelligence, academic achievement, expressive and receptive language, all aspects of memory, attention/concentration, concept formation and abstract reasoning, bilateral upper extremity motor functions, emotional functions, as well as effort and motivation.

With respect to memory assessment, the TUMC database includes data from evaluations using the Wechsler Memory Scale – Third Edition (WMS-III), and the California Verbal Learning Test – Second Edition (CVLT-II). The CVLT-II, as noted previously, is used as a measure of the patient's ability to acquire, retain, recall and recognize verbal information both immediately and after a time delay. The word lists used in the CVLT and CVLT-II can be organized into four semantically related lists (i.e. vegetables, tools, items of clothing, modes of transportation) that, theoretically, provide a structure that facilitates recall of individual list items. The recent (2000) revision of the CVLT is distinguished from the original version by the inclusion of several new subtests, including the Delayed Recognition subtest examining recognition memory for List A items. In this subtest, the patient is read a list of 32 words, consisting of List A items, List B items, and foils, some of which are semantically related to the target words. The test requires the patient to respond affirmatively to only the List A words. The Delayed Forced Choice subtest, which is administered at the end of the test, the patient is read a list of 16 word pairs, with each pair containing a target word from List A and a foil. The patient must identify the word that was presented in List A, and the patient's performance is expressed as the percent of List A words that are identified correctly. The words in each pair are further distinguished by their level of complexity, with the List A words being relatively concrete nouns (i.e. "onion", "cow", "desk") whereas the foils are more abstract nouns that are encountered with less frequency in everyday speech. (i.e. "majority", "technique", "sprinkler"). Connor et al. (as cited in Delis, Kramer, Kaplan & Ober, 2000) tested this format in a population of traumatically brain-injured patients, as well as a population of subjects exhibiting suboptimal effort on the Hiscock forced choice test. The results revealed that this forced choice format had 80 percent sensitivity and 97 percent specificity in correctly identifying patients attempting to feign memory deficits. The CVLT-II normative data indicate that more than 90 percent of the entire normative sample scored perfectly on this forced choice subtest (Delis, et al., 2000, p. 162). These very high ceiling effects demonstrate the utility of this subtest as a measure of effort, as the overwhelming majority of patients, even those with significant cognitive impairments, are able to complete this subtest without error.

The WMS-III is a comprehensive battery of tests designed to measure immediate and delayed memory in both the verbal and visual domains. Eight index scores are calculated from the test data, including the Auditory Recognition Delayed Index (ARD), which provides a measure of the patient's ability to discriminate target verbal information from foils, and represents the combination of two of the verbal subtests found on the WMS-III (Logical Memory-II Recognition and Verbal Paired Associates-II Recognition). The Recognition portion of the Visual Reproduction-II subtest is used to examine the patient's recognition memory for non-verbal materials, in this case, a set of five geometric designs of increasing complexity. These scores (The Auditory Recognition Delayed Index; hereafter referred to as ARD, and the Visual Reproduction-II Recognition subtest) were of particular interest in this study, as they provided measures of recognition memory, and were included in the statistical analyses detailed in the following section.

In addition to these measures of cognitive functioning, data representing the results of comprehensive psychological evaluation for the identified subject pool were taken from the database, specifically, the available F, Hs and Hy scale T score from the MMPI-2, and the total scores from the Beck Depression Inventory – Second Edition (BDI-II) and the Beck Anxiety Inventory (BAI).

The Neuropsychological Symptom Checklist (NSC; Schinka, 1983) is also a self-report measure of current and premorbid problems experienced by the patient. The NCS is comprised of 93 items, which can be endorsed by the respondent with a check mark. Although the test items are not listed in specified groups, they pertain to possible problems within such domains as sensory and motor functioning, pain, seizure and/or syncopal activity, disorientation, memory problems, unusual experiences (visual/auditory/tactile hallucinations), information processing, attention/concentration, judgement, praxis, language, mood, behavior, the presence of other medical problems, surgical history, substance use, chemical exposure and family medical history. The checklist was developed as a means of identifying patient complaints with respect to physical, cognitive and emotional functioning; however, the author notes that no scoring system exists by which to quantify patient responses (personal communication, July 2, 2002). However, the simple yes/no response format of this inventory allows for tabulation of the frequency of item endorsement within the subject sample. For the purposes of the current study, the frequencies of subjective complaints of memory problems were calculated from subject responses to the NSC.

The Tulane University test battery also included measures of symptom validity, such as the Test of Memory Malingering (TOMM; Tombaugh, 1996), the Word Memory Test (WMT; Green et al., 1996), the Rey 15-Item Test and the forced-choice subtest from the California Verbal Learning Test – Second Edition (CVLT-II; Delis et al., 2000). The format of each of these tests has been described previously in this paper. Subject performance on each of these tests is scored with respect to empirically derived cutoff scores; with subjects scoring below these cutoff scores diagnosed as probable malingerers.

1.9.3. Analysis

The comprehensive nature of the Tulane Neuropsychology Laboratory database, and the flexible nature of the coding system used to store the data, allowed for data to be extracted in a number of ways depending on the nature of the specific analysis. In order to investigate the hypotheses in which the likelihood of subject malingering was a central factor, subjects were divided into groups based on a subjective rating of malingering probability. The criteria for such determinations were applied by F. William Black, Ph.D., Director of the Neuropsychology Laboratory at Tulane University, following the recommendations of Slick et al. (1999) as noted previously. The specific criteria for the determination of malingering probability are presented in detail in Appendix A. Dr. Black comprehensively reviewed all subject files in the absence of any information that would allow for patient identification. Following these subjective rating criteria, all subjects were assigned one of the following ratings: 1 – Absolute Malingering, 2 – Probable Malingering, 3 – Possible Malingering and 4- Not Malingering. Accordingly, four groups of patients were available for study, according to this coding system. In addition, data corresponding to referral source were also available, and included those subjects referred by defense and plaintiff attorneys, Workman's Compensation, physicians and/or other psychologists, by individuals requesting psychoeducational data, and other referrals that did not fall into one of these pre-specified categories. This coding system allowed for the examination of any effects related to referral source.

In order to examine the possible change in symptom validity test performance through time, archived data were extracted from the available database and coded according to year, spanning the years 1998 through 2002. Data collected in the year 2003 were also available, but did not encompass the entire calendar year. In addition, data pertaining to age, sex, level of education, number of previous neuropsychological evaluations, and previous involvement in litigation were also extracted for each of the cases used for analysis, in order to describe the demographic characteristics of the subject samples used in the study. No demographic differences were expected to exist among the subject groups.

With respect to the composition of the groups used in this study, a potential confound existed regarding the neurological status of the subjects within the database. More specifically, it might have been be expected that any patients with brain impairment would score below their neurologically intact counterparts on specific tests, regardless of litigation status, especially on measures of memory, which tend to be sensitive to cognitive impairment. Data pertaining to the neurological status of each subject was available in the database, thus providing the ability to control for the effects of acquired brain impairment in the analyses. Interference from cognitive impairment was not expected to interfere with the analyses proposed for three of the malingering measures (the TOMM, WMT and CVLT-2 Forced Choice), as the normative data for these tests demonstrates that they are robust to the effects of even moderate to severe dementia. Previous literature has suggested that the 15 Item Test may demonstrate false positive results when used with moderately to severely mentally retarded individuals. As none of the subjects in the available database carried such diagnoses (moderate to severe dementia or mental retardation), no adverse effects on test performance that could be attributable to such diagnostic categories were expected.

In order to conduct the proposed analyses, adequately sized samples were necessary. Accordingly, all available data for the following tests were extracted from the database: the Delayed Yes/No Recognition and the Delayed Forced Choice Recognition trials of the CVLT-II, the Auditory Recognition Delayed Index from the WMS-III, the Delayed Recognition scores from the Wechsler Memory Scale Visual Reproduction subtest, the T score for the F, Hs and Hy scales of the MMPI-2, the scores for each of the three trials of the TOMM (Trial 1, Trial 2 and Recognition), the scores for each trial of the WMT (Immediate Recognition, Delayed Recognition and Consistency), the total number of items correctly recalled for the Rey 15 Item Test, the total scores for the BDI and BAI, and the responses to the memory problems question on the NSC. In order to conduct statistical analyses of the data enumerated above, it was necessary to export the stored data into an appropriate statistical analysis program. Due to incompatibilities between the data storage and analysis programs used in the Tulane Neuropsychology Laboratory, data could not be transferred directly from the Access '97 database into SPSS 11.5, which was used for all data analysis procedures. Therefore, an interim step, in which all data to be analyzed were exported from the Access '97 to Excel '97, initially completed. The data were then transferred from the Excel '97 file into SPSS Version 11.5.

As noted previously, the composition of the neuropsychological assessment battery used in the Tulane University Neuropsychology Laboratory has changed over time, as both new and revised tests have become available. In addition, the composition of the test battery often differed with respect to the referral source, such that most non-litigating subjects were not typically administered more than one test of symptom validity, if any. This practice has recently changed in the Neuropsychology Laboratory, due to growing interest in the phenomenon of malingering or symptom exaggeration in non-litigating patients, and to provide increased continuity in the data set. However, for the purposes of the present study, insufficient sample sizes were available to examine such effects and, therefore, only data from litigating patients were available to test those hypotheses regarding symptom validity test performance. In addition, some inconsistencies were seen in the numbers of specific cognitive tests administered, again due to changes in the composition of the standard test battery. While such inconsistencies may be seen as a weakness in the design of this study, it should be emphasized that the evaluations conducted in the Neuropsychology Laboratory are primarily clinical in nature, and the composition of the test battery is constructed accordingly. This practice is supported by Standard 9.08 (b) of the Ethical Standards of the American Psychological Association (APA, 2002), which mandates that "Psychologists do not base such decisions or recommendations on tests and measures that are obsolete and not useful for the current purpose", indicating that the most recent version of a given test be used in the absence of justifiable reasons to continue to use an earlier version. As an example, given that the most recent version of the CVLT provides improved normative data the inclusion of new subtests, this test was added to the Neuropsychology Laboratory battery when it became commercially available. To account for the inequalities in sample sizes, individual samples were extracted from the database for each of the hypotheses to be tested, and include data for the test(s) of interest as well as the demographic data pertaining to that particular sample. This method was used to avoid the need to remove subjects who did not have data pertaining to each test under consideration in this study, as doing so would have significantly reduced the number of data points available for analysis had all data been compiled into a single data set. In order to examine the effectiveness of specific symptom validity tests over time, data for the tests of interest outlined in the above hypotheses were coded according to year (1999, 2000, 2001, 2002 and 2003).

Analysis of subjective memory complaints in litigating versus non-litigating patients consisted of a simple tabulation of the frequency with which such complaints were made on the NSC. After tabulating the number of such complaints in each group, a

Mann-Whitney U test was calculated to determine if the groups were significantly different with respect to subjective memory complaints.

2: Results

Preliminary examination of the data of interest demonstrated notably non-normal distributions in all cases. In addition, the study's use of categorical data required the use of nonparametric statistical procedures. Given these considerations, the nonparametric equivalent of the one-way analysis of variance test for independent groups, the Kruskal-Wallis procedure, was used in all cases where a significance test for *k* independent groups was needed. According to Andrews (1954, as cited in Siegel, 1956), the Kruskal-Wallis procedure has an efficiency of $3/\pi = 95.5$ percent, making it a relatively powerful procedure for nonparametric analysis of variance in the *k*-group case. As *post hoc* tests are not available for nonparametric analysis of variance, it was necessary to conduct multiple Mann-Whitney U tests, the nonparametric equivalent of the *t* test. Again, according to Siegel (1956, p. 116), the Mann-Whitney U test is "one of the most powerful of the nonparametric tests" and is "a most useful alternative to the parametric *t* test, when the researcher wishes to avoid the *t* test's assumptions, or when the measurement in the research is weaker than interval scaling."

Given the need for multiple (72) univariate comparisons in the course of the data analysis, Bonferroni correction to control for experimentwise Type 1 error was required. This correction, for a nominal alpha of .05, yielded a very restrictive critical value of .0001. However, given that this study was, essentially, comprised of a number of smaller studies with distinct samples, separate correction factors were calculated for each of these respective analyses. This step was undertaken as there were concerns that the overly restrictive experimentwise alpha level would obscure any findings of interest within these smaller, theoretically discrete analyses. Where indicated, the results of both the restrictive and less restrictive approaches to the data analysis are presented and the results are interpreted for each approach.

2.1. General Demographics of the Sample

Demographic data for the entire subject sample are presented in Table 1. Examination of these data reveal that the overall sample was predominantly male (64.1%) and that approximately two-thirds of the subjects had previously undergone at least one comprehensive neuropsychological assessment. Defense counsel had referred the majority of these subjects (50.4%). Although approximately 34 % of the sample denied having been involved in previous litigation, this information was lacking for approximately 36% of the total sample.

Table 1. Demographic Data for the Total Sample

Gender N = 1290	Male 831 (64.1%)	Female 459 (35.4%)			
Number of previous neuropsych. evaluations N = 914*	No evals. 354 (27.3%)	One eval. 424 (32.7%)	Two evals. 96 (7.4%)	Three or more evals. 25 (1.9%)	
Previous Litigation N = 1274	No 436 (33.6%)	Yes 142 (11%)	DK** 463 (35.7%)		
Referral Source N = 1290	Defense 653 (50.4%)	Plaintiff 328 (25.3%)	Workman's Comp. 165 (12.7%)	Other^ 144 (11%)	
Malingering Rating N = 979*	Absolute 190 (14.7%)	Probable 147 (11.3%)	Possible 149 (11.5%)	No 493 (38%)	
Year of Administration N = 1290	1999 159 (12.3%)	2000 223 (17.2 %)	2001 376 (29%)	2002 492 (38%)	2003 40 (3.1%)

* This information has not been systematically collected over time; therefore, some data are missing.
** DK (don't know) = data were either not available or not reported.
^ Includes medical, psychological, psychoeducational, and other evaluations.

Since the overall study was, in effect, comprised of multiple, but essentially distinct, studies, separate demographic data were compiled for each of these samples. These data are presented in Table 2.

Table 2. Demographic Data for Subject Samples by Tests Administered

	Rey 15 Item	TOMM	WMT	CVLT-2 Y/N Hits	MMPI-2 BAI/BDI	WMS-III ARD	WMS-III VR Rec.	NSC
Education (mean yrs.)	12.8	12.8	12.9	12.8	13.4	13.6	12.7	14.9
Gender Male (%)	66.3	65.5	62.7	66.7	58.5	66.3	69.6	62.9
Female (%)	33.7	34.5	32.8	33.3	41.5	33.7	30.4	37.1
Age (mean yrs)	39.1	38.6	39.3	39	37.9	40.5	40.5	40.6
Mos. Post Injury (mean)	31.7	38.5	38.6	36	36.4	32.3	29.5	36.3
Number of Previous Neuropsy. Evals. (%)								
0	16	26.6	31.3	31.7	21.6	44.6	47	49.3
1	23	34.8	47.8	50.8	19.5	43.4	40.9	36.1
2	7	7.9	12.7	12.7	3	7.2	7.8	9.3
≥ 3	1.2	1.8	2.2	3.2	.9	3.6	3.5	2
Previous Litigation (%)								
No	31.7	30.7	24.6	30.2	44.8	26.5	32.2	16.6
Yes	9.9	9.7	10.4	15.9	10.7	13.3	10.4	12.7
DK	58.4	43.8	60.4	54	.3	59	56.5	70.2
Referral Source (%)								
Defense	54.7	53.6	60.4	67.9	39.6	43.4	45.2	49.3
Plaintiff	30	31.5	23.1	20.6	25.6	15.7	14.8	16.6
WC	15.2	15	11.9	17.5	7.9	13.3	11.3	8.8
Other	NA	NA	NA	NA	26.8	26.7	28.7	30.7
Maling. Rating (%)								
Absolute	14.4	13.5	20.9	28.6	4.6	21.7	20	12.7
Probable	11.9	10.9	13.4	17.5	.7	13.3	13	10.2
Possible	10.7	10.5	14.2	19	6.7	18.1	13	13.7
No	37	37.8	47.8	34.9	28.4	47	53.9	63.4
Year of Testing (%)								
1999	20.6	21	None	None	NA	NA	NA	NA
2000	46.9	24	None	None	NA	NA	NA	NA
2001	32.5	30	47.8	27	NA	NA	NA	NA
2002	None	52.1	47.8	73	NA	NA	NA	NA
2003	None	None	None	None	NA	NA	NA	NA

Examination of these data indicates that significantly more men were evaluated in the Neuropsychology Laboratory than women for the years 1999 through part of 2003 ($z = 14.9$, $p = .000$). It is interesting to note that defense counsel referred the majority of the subjects seen for evaluation (50.4%), which is generally seen as a rather contentious and unnecessary maneuver by many subjects pursuing personal injury litigation, as most of these individuals have, at that point, already undergone a comprehensive neuropsychological examination at the behest of their own attorney. Given the nature of this clinical situation, in which it may be expected that subjects may be motivated to prove their cognitive impairment to the "defense neuropsychologist", only 14.7 percent of the total sample was rated as Absolute on the Malingering Rating Scale. In addition, only 36 percent of the sample received any rating indicative of some degree of malingering, less than the proportion of the sample that was rated as not malingering.

2.2: Symptom Validity Test Performance over Time

2.2.1: Rey 15 Item Test

Differences in the number of omission errors over the three years of collected data (1999, 2000l, 2001) are displayed in Figure 1. Inspection of this figure suggests that, contrary to the hypothesized effect, more subjects appeared to make greater numbers of omission errors over time.

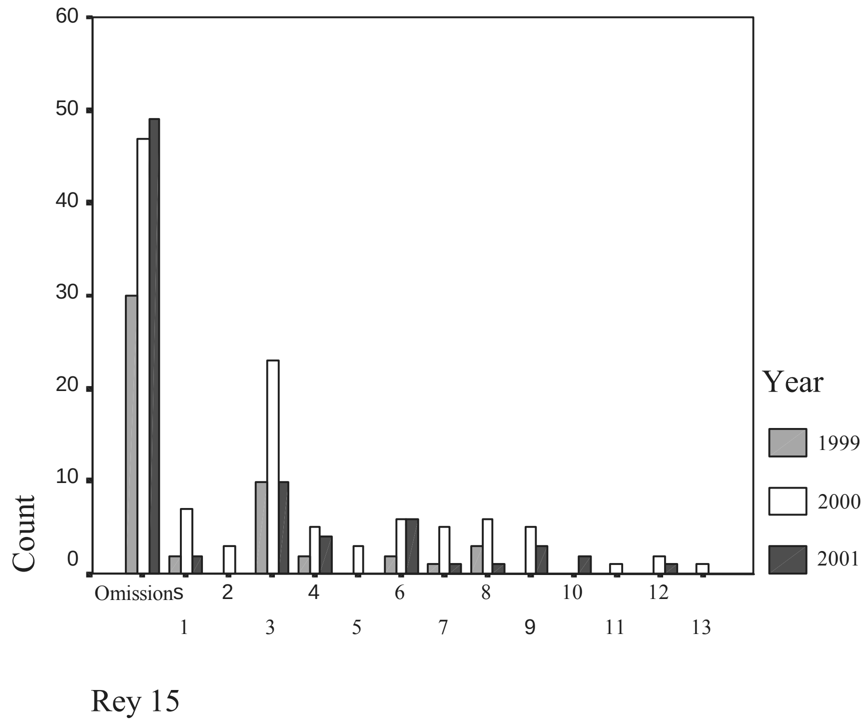

Figure 1. Rey 15 Item Test Scores by Year of Administration

It was hypothesized that overall performance on the Rey 15 Item test would become more normal over time, that is, that the majority of test scores would be within the normal range of nine correct responses and, therefore, not indicative of malingering or reduced effort. To investigate this hypothesis, a Kruskal-Wallis test was conducted to evaluate differences among the three consecutive years of data in consideration (1999, 2000 and 2001). The test, which was corrected for tied ranks, was not significant, χ^2 (2, N = 243) = 8.23, p = .01, based on the experimentwise alpha of .001. Accordingly, follow-up Mann Whitney U tests were not performed.

As noted previously, significant restrictions were placed on the experimentwise alpha level, due to the very large (72) number of pairwise comparisons. As a result, the corrected alpha level was set at a very restrictive level of .001. While any attempt to re-adjust this alpha level may appear to be an attempt to "fish" for significant results, it must be reiterated that each of the hypotheses in this study were based on discrete samples that, in effect, serve as individual studies in their own right. With that in mind, it was felt that a re-examination of these data as a discrete study was warranted. The results of this re-analysis revealed a significant Kruskal-Wallis test, with χ^2 ($df = 2$, $N = 243$) = 8.23, p = .016, based on the standard nominal alpha level of .05. A very strong trend toward significance was found only between the years 2000 and 2001, with $z = -2.35$, $p = .019$, based on less restrictive corrected alpha level of .017. Subjects in the year 2000 averaged 3 omission errors compared to an average of 2 such errors in the year 2001. These data suggest that, as a group, subjects taking this test in 2001 made relatively fewer errors than those tested in the previous year, thus providing some support for the proposed hypothesis that scores would normalize over time. These data are consistent with those of Grieffenstein & Baker (2002), who found a similar trend in their Rey 15 Item test results over a 10-year period.

2.2.2: The Test of Memory Malingering

Analyses were conducted on data pertaining to Trials 1 and 2 of the TOMM, which are both immediate recognition trials. Data pertaining to Trial 3 (Delayed Recognition) was omitted from analysis, as this optional trial was not consistently administered in the Neuropsychology Laboratory, resulting in missing data. For each of the two immediate recognition trials of the TOMM, Kruskal-Wallis analysis of variance procedures were conducted to determine the existence of differences between scores on these trials according to the year of administration (1999, 2001, 2002 and 2003). The results of the Kruskal-Wallis procedure for the TOMM Trial 1 was not significant, where χ^2 (df = 3, N = 267) = 6.36, p = .095. Similarly, the results were not significant for the TOMM Trial 2 data, where χ^2 (df = 3, N = 267) = 5.41, p = .14. These results were also not significant when a standard nominal alpha level of .05 was applied; therefore, the lack of significant findings does not appear to be due to the restrictive nature of the stringent experimentwise alpha of .001 employed in this study. Follow-up Mann-Whitney U tests could not be conducted, due to the lack of significant findings in the omnibus significance tests. The distributions of these data are represented in Figures 2 and 3.

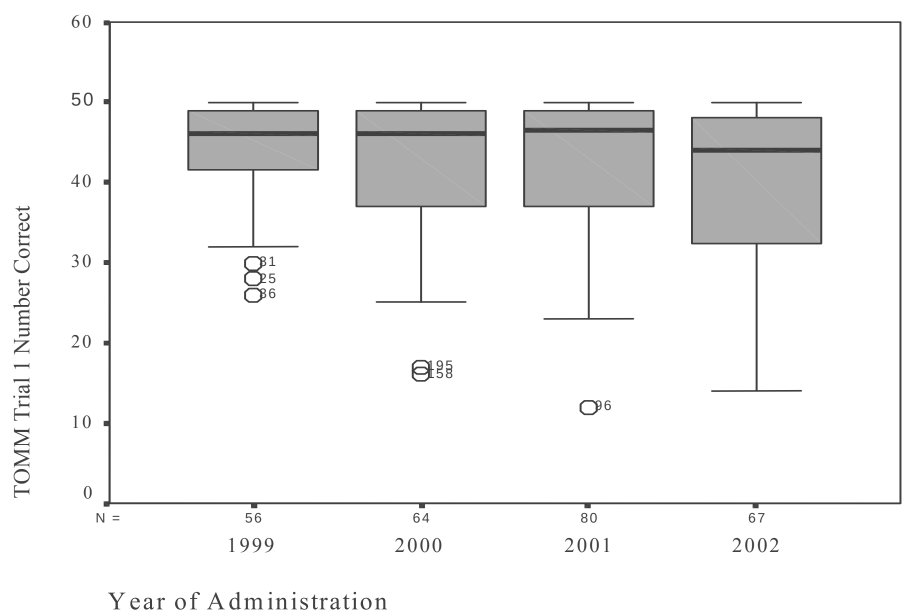

Figure 2. Distribution of TOMM Trial 1 Scores for Years 1999 Through 2002

Although a significant change in TOMM scores was not found over time for either Trial 1 or Trial 2, significant differences were seen when TOMM scores were compared across levels of the Malingering Rating Scale (1 = absolute, 2 = probable, 3 = possible and 4 = no). The results of a Kruskal-Wallis test to evaluate the relationship of TOMM Trial 1 scores among the four levels of Malingering Rating scores was significant, where χ^2 (df = 3, N = 194) = 82.03, p = .000. Follow-up Mann-Whitney U tests were then conducted to evaluate the presence of significant differences between levels of the Malingering Rating Scale. The results are presented in Table 4.

Table 3. Pairwise Comparisons of Malingering Rating Scores for the TOMM Trial 1

Malingering Rating Scale	Absolute	Probable	Possible	No
Absolute		z = -2.65 p = .008	z = -5.71 p = .000*	z = -7.89 p = .000*
Probable			z = -3.22 p = .001*	z = -5.26 p = .000*
Possible				z = -2.53 p = .011

* $p \leq .001$

According to these results, significant differences exist between most of the Malingering Rating groups, namely, between Absolute and Possible malingerers, Possible malingerers and those falling into the No group, and those in the Probable versus the No group. The difference between the Possible and Probable malingerers was just within the range of significance, suggesting that some unknown but notable differences may exist between subjects in these two groups, or in the criteria used to make the malingering ratings.

Similar results were found when a Kruskal-Wallis test was conducted to evaluate differences between the Malingering Rating scores and subject performance on the TOMM Trial 2. The omnibus test was significant, χ^2 (df = 3, N = 194) = 94.9, p = .000, again despite the restrictive nature of the experimentwise α level of .001. Follow-up Mann-Whitney U tests were then conducted to evaluate the presence of significant differences between levels of the Malingering Rating Scale. The results are presented in Table 5, and the distributions of these data are presented in Figure 3.

Table 4. Pairwise Comparisons of Malingering Rating Scores for the TOMM Trial 2

Malingering Rating Scale	Absolute	Probable	Possible	No
Absolute		z = -3.21 p = .001*	z = -5.70 p = .000*	z = -8.75 p = .000*
Probable			z = -3.10 p = .002	z = -6.25 p = .000*
Possible				z = -3.48 p = .000*

* $p \leq .001$

These results demonstrate significant differences between all Malingering Rating groups on Trial 2 of the TOMM, with the exception of the Probable versus Possible malingerers.

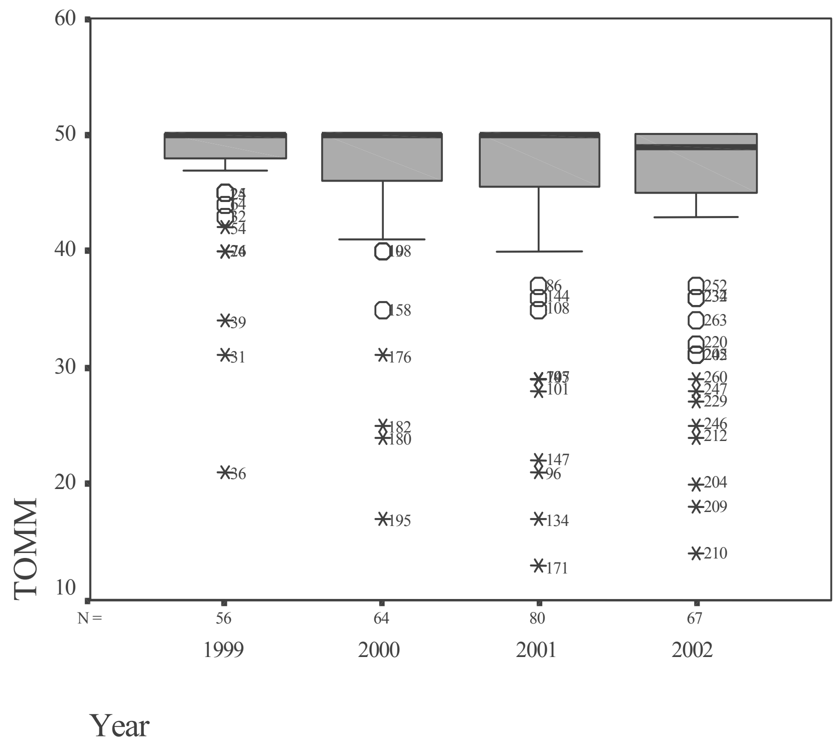

Figure 3. Distribution of TOMM Trial 2 Scores for Years 1999 through 2000

2.2.3: Word Memory Test

Data pertaining to the two years (2001 and 2002) in which the WMT was administered in the Neuropsychology Laboratory were analyzed with Mann-Whitney U tests, to evaluate the hypothesis that the WMT scores would, on the average, become more normal over time and, therefore, indicate a lower relative incidence of malingering over time. As with the other symptom validity tests evaluated here, the expected normalization in WMT scores over time was hypothesized to be related to subject familiarity with the test, or at least its premise, via the provision of such information by attorneys or from Internet sites or other sources devoted to the dissemination of such information. The results of the Mann-Whitney U tests were not significant for any of the three trials of the WMT. Specifically, an analysis of Immediate Recognition scores revealed $z = -2.06$, $p = .013$. For Delayed Recognition, $z = -2.06$, $p = .039$ and for Consistency, $z = -2.46$, $p = .014$.

Re-analysis of these results as a discrete study, with a corresponding nominal alpha level of .017 indicates that a significant difference exists between years of assessment for both the Immediate Recognition and Consistency scores, with subjects in the second year of assessment (2002) scoring below those assessed in 2001 on all three trials of this test. Overall, Immediate Recognition scores were lower than Delayed Recognition scores in both groups; therefore, the significant findings found in the Immediate Recognition and Delayed trials are likely related, since discrepant scores on the Immediate and Delayed Recognition trials produce low Consistency scores. Overall, these findings suggest a decline in WMT scores over time, in direct contrast to the hypothesis that such scores would become less aberrant over time.

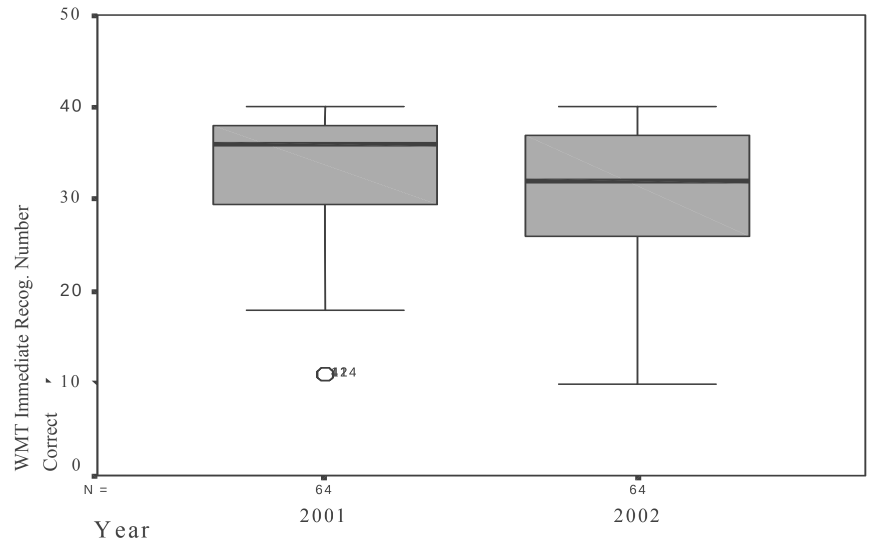

Figure 4. Distribution of Word Memory Test Immediate Recognition Scores for the Years 2001 through 2002

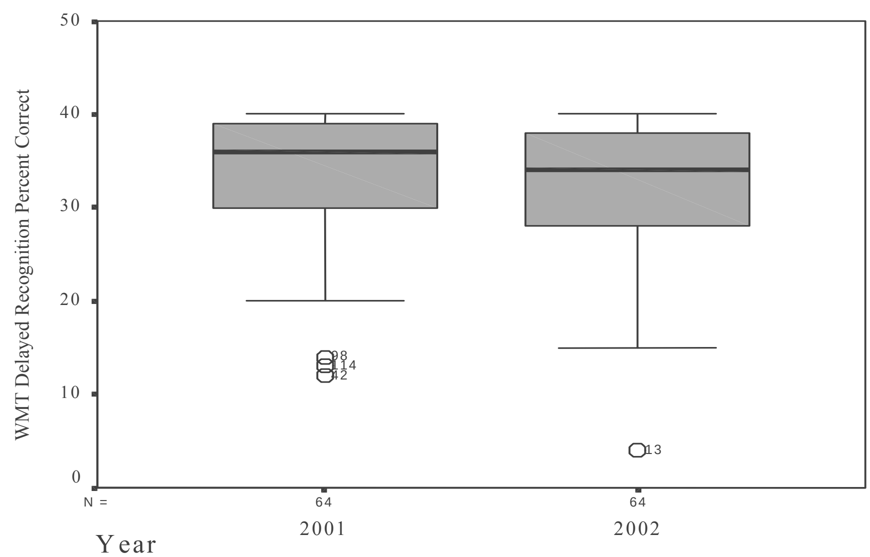

Figure 5. Distribution of Word Memory Test Delayed Recognition Scores for the Years 2001 through 2002

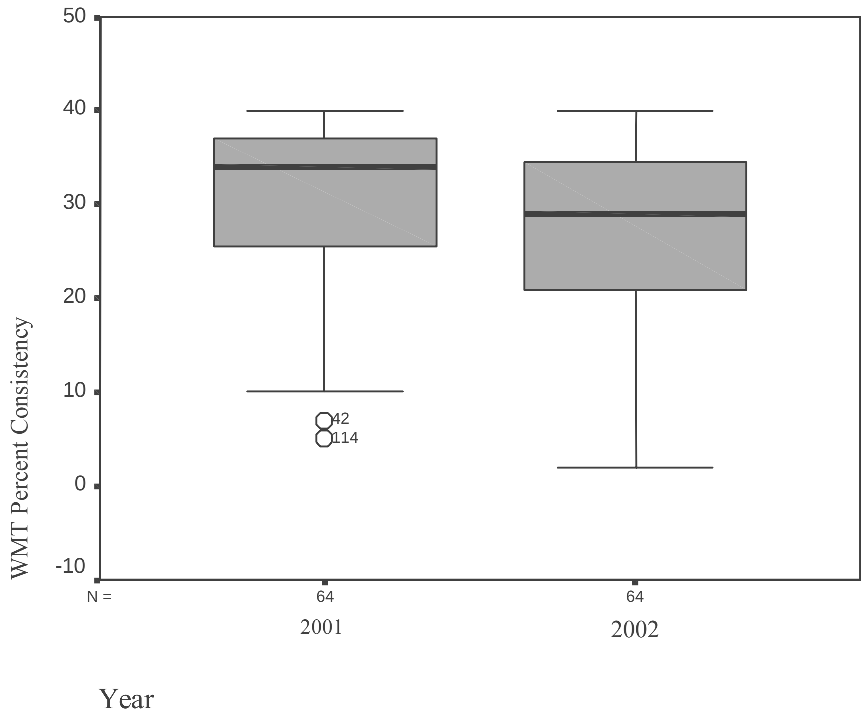

Figure 6. Distribution of Word Memory Test Consistency Scores for the Years 2001 through 2002

2.2.4: The California Verbal Learning Test-2 Yes/No Hits Subtest

When viewed graphically (Figure 2), a wider variance in test scores is apparent in 2002 compared to 2001, with a greater number of abnormal test scores seen in 2002, contrary to the hypothesized effect of an overall normalization of scores over time.

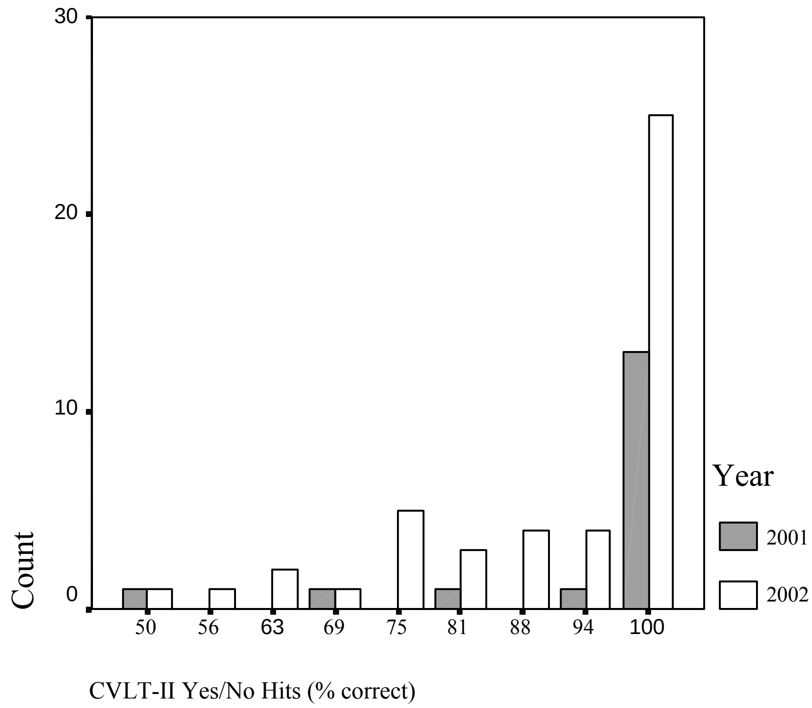

Figure 7. Distribution of CVLT-II Yes/No Hits Scores for the Years 2001 through 2002

This observation is supported by the results of the Mann-Whitney U test, which was not significant, $z = -1.4$, $p = .16$, irrespective of the use of either the restrictive experimentwise alpha level of .0001 or a standard alpha of .05. Therefore, there does not appear to be a trend toward normal scores on CVLT-II Yes/No subtest over time.

2.3: Performance on the MMPI-2 Validity and Clinical Scales (F, Hs, D, Hy, Pt), Beck Depression Inventory – Second Edition, and Beck Anxiety Inventory as a Function of Malingering Rating

A Kruskal-Wallis test was conducted on each of the target MMPI-2 scales (F, Hs, D, Hy and PT) independently, to assess its relationship, if any, to the Malingering Rating Scale. The results of these tests were significant for Hs, χ^2 (df=3, N = 153) = 27.44, p =

.000 and for Hy, χ^2 (df=3, N = 153) = 20.25, p = .000. The Kruskal-Wallis test for scale Pt was nearly significant, χ^2 (df=3, N = 153) = 14.71, p = .002, but was not significant for the D scale, χ^2 (df=3, N = 153) = 13.26, p = .004. A Kruskal-Wallis conducted to evaluate scores on the Beck Depression Inventory, Second Edition (BDI-II) as a function of Malingering Rating was not significant, χ^2 (df=3, N = 153) = 13.21, p = .004; however, the same test conducted with the Beck Anxiety Inventory (BAI) was significant, χ^2 (df=3, N = 153) = 16.34, p = .001. Follow-up Mann-Whitney U tests were conducted to determine which pairs of Malingering Rating scores were significantly different for the Hs and Hy scales and the BAI, based on Malingering Rating scores. Significant differences were found only between the Absolute and No ratings for the Hs, z = -4.83, p = .000 and the Hy, z = -4.23, p = .000, scales, and the BAI, z = -3.62, p = .000.

2.4: Evaluation of Verbal Memory Test Performance as a Function of Malingering Rating

It was hypothesized that subject performance would vary on these verbal subtests depending on the subject's rating on the Malingering Rating, with subjects rated as Malingering generating scores below the accepted cutoff for either, or both, subtests. These hypotheses were supported by Kruskal-Wallis tests, which were significant for both the WMS-III Auditory Recognition Delayed subtest, where χ^2 (df=3, N = 83) = 24.02, p = .000, and the California Verbal Learning Test – Second Edition (CVLT-II) Yes/No Hits subtest, where χ^2 (df=3, N = 153) = 28.34, p = .000. Follow-up Mann-Whitney U tests demonstrated significant differences between Malingering Ratings of Absolute and No (WMS-III, z = -4.5, p = .000; CVLT-II, z = -4.44, p = .000). When non-litigating subjects were removed from the data analysis, these significant differences were not found for either the WMS-III Auditory Recognition Subtest, where χ^2 (df = 2, N = 83) = .48, p = .79 or the CVLT-II Yes/No Hits subtest, where χ^2 (df = 2, N = 83) = .59, p = .74, suggesting that malingering ratings did not differ between the three groups of litigating subjects (defense-referred, plaintiff-referred, or Workman's Compensation).

Considering the WMS-III ARD Index data as a discrete data set, the experimentwise alpha was set at .008, less restrictive than the .0001 for the entire study. At this less stringent level, significant differences were found on Mann Whitney U tests comparing Absolute and Possible Malingerers, where z = -2.86, p = .004. This finding is in addition to the significant difference, noted above, found between absolute malingerers and those rated as not malingering. Therefore, this subtest seems to provide some incremental information regarding the validity of the responses generated by subjects who were believed to be making some attempt to feign memory impairment.

Similar re-analysis of the CVLT-II Yes/No Hits data, with the same adjusted alpha level of .008, revealed significant differences between Absolute and Possible malingerers, where z = -2.99, p = .003, and Probable malingerers and those rated as not malingering, where z = -3.49, p = .000. Overall, these results suggest that notably aberrant scores on these subtests provide additional evidence of feigned verbal memory impairment. The distributions of the WMS-III Auditory Recognition and CVLT-2 Yes/No Hits are depicted in Figures 8 and 9, respectively.

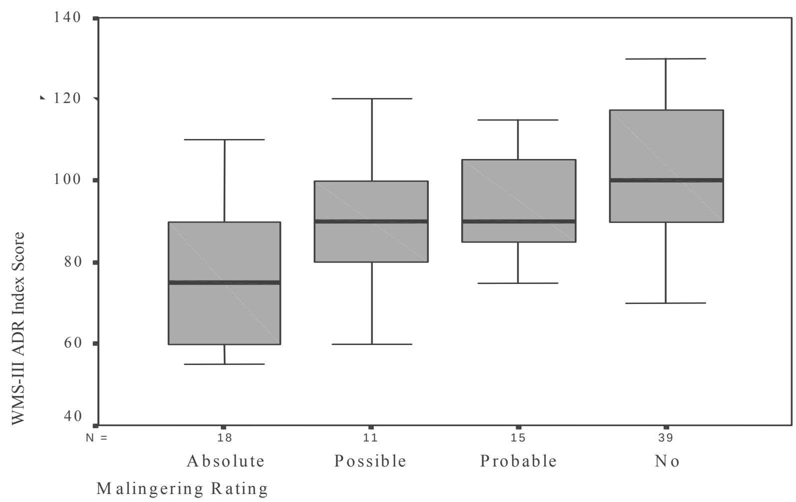

Figure 8. Distribution of WMS-III Auditory Recognition Delayed Index Scores by Malingering Ratings

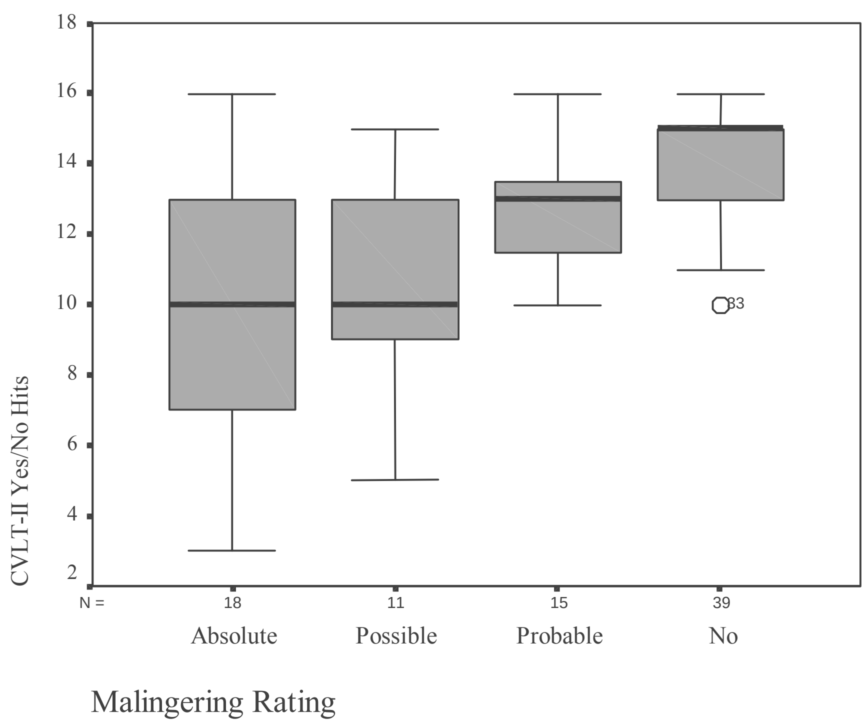

Figure 9. Distribution of CVLT-II Yes/No Hits by Malingering Rating

2.5: Evaluation of Visual Reproduction Recognition Memory Test Performance as a Function of Malingering Rating

Again, the Kruskal-Wallis test was conducted to evaluate differences in WMS-III Visual Reproduction Recognition in relation to Malingering Rating Scale scores, with the same expectation as detailed in the previous section pertaining to verbal memory test data. The overall result was significant, χ^2 (df=3, N = 115) = 16.64, p = .000. Follow up Mann-Whitney U tests demonstrated a significant difference between subjects rated Absolute and No on the Malingering Rating Scale. However, no significant differences were found when non-litigating subjects were removed from the analysis χ^2 (df=2, N = 82) = .22, p = .89.

Fewer significant differences were found between malingering groups for this data set compared to those found in the verbal memory data described in the previous section. Again, the adjusted alpha level of .008 was applied, resulting in a significant difference between only the Absolute malingerers and those rated as not malingering, where z = -3.69, p = .000. The distribution of WMS-III Visual Reproduction Recognition scores is found in Figure 10 on the following page.

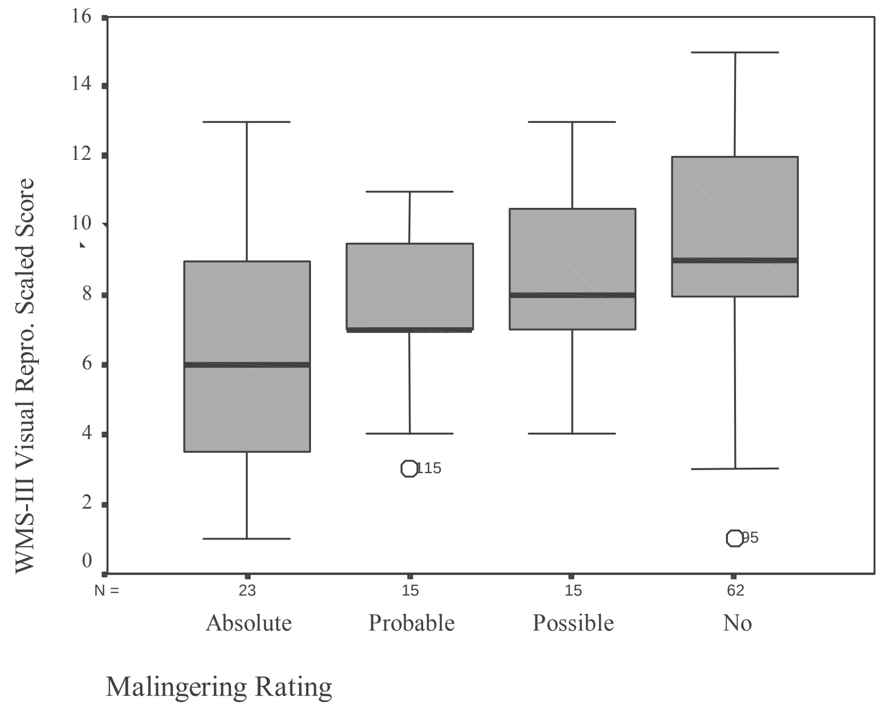

Figure 10. Distribution of WMS-III Visual Reproduction Recognition Scores by Malingering Rating

2.6: Neuropsychological Symptom Checklist Memory Problems Scores as a Function of Malingering Rating

A Kruskal-Wallis test to determine significant differences between self-report of memory problems on the NSC as a function of Malingering Rating was not significant, where χ^2 (df=3, N = 204) = 1.87, p = .59.Given the non-significant findings, follow-up tests were not performed. These results are unchanged when these data are analyzed as a discrete set and a nominal alpha level of .05 is applied.

When displayed graphically, as in Figure 11, it would appear that the rate of subjective memory complaints differs dramatically between referral source groups, with a very large number of defense-referred subjects making memory complaints, compared to a relatively small number of such complaints made by subjects referred for psychoeducational or other, non-medically oriented evaluations. Despite these apparent discrepancies, the omnibus significance test between these groups was non-significant, where χ^2 ($df = 5$, $N = 205$) = 8.4, $p = .14$, suggesting that the rate of subjective memory complaints was similar across all referral groups. The distribution of the NSC Checklist responses is found in Figure 11, on the following page.

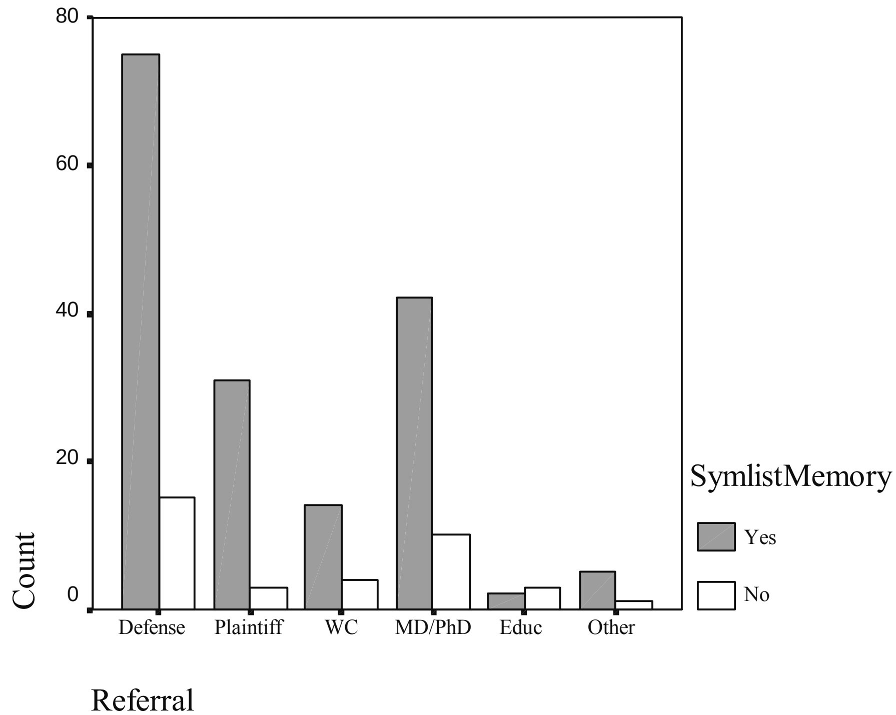

Figure 11. Distribution of Neuropsychological Symptom Checklist Memory Problems Scores by Malingering Rating

3: Discussion

The present study examined the proposed hypothesis that, as a whole, subjects' performances on specific symptom validity tests would demonstrate a trend toward more normal scores through time, corresponding to fewer cases in which such test scores would be classified as indicative of malingering. Only one study to date has been undertaken to examine this proposed effect (Greiffenstein & Baker, 2002), and then only with the Rey 15 Item test. The results of that study, although hampered to some degree by small sample sizes, demonstrated changes in Rey 15 Item test performance over time, with more normal scores obtained in later years; however, these results failed to reach significance. Therefore, while it appears that information pertaining to the "philosophy" of symptom validity testing, ways to "beat" the tests, and provision of specific test data are available to any person with access to the Internet (Ruiz, et al., 2002), as well as suggestions that a majority of attorneys feel obligated to inform their clients about the use and content of such tests (Wetter & Corrigan, 1995), empirical investigation of such effects have not been undertaken. The reason for this is unclear, but may be related to lack of data, due either to insufficient sample sizes (in terms of litigating subjects) or inconsistent use of specific tests over time (due to the aforementioned lack of litigating subjects and/or changes in test battery composition). The present study was designed to take advantage of a large database in a hospital-based clinical practice, in which approximately 50 percent of the neuropsychological evaluations conducted are done so in the context of civil litigation cases in which permanent cognitive disability is alleged. Symptom validity testing, and the detection of malingering, dissimulation or exaggeration, was an integral part of the overall clinical approach in this practice, which emphasized the validation of all aspects of the data collected for analysis. Therefore, a relatively large body of data pertaining to the evaluation of symptom validity was available, as well as to various aspects of functioning (intellectual, cognitive, academic, physical and emotional).

As noted previously, the structure of this study and the nature of the data used allowed for the formation of multiple related, yet independent, hypotheses. Given that the investigation of each hypothesis could serve as a unique study, it was felt that a re-analysis of the data collected to test individual hypotheses was warranted. Doing so would lessen the notable restrictions in interpretation imposed by the experimentwise alpha level of .0001. While the argument can certainly be made that any re-analysis of the data constitutes an attempt to find significance where none exists, the alternative argument lies, again, in the overall construction of the study, in which multiple, discrete studies were combined to evaluate an important set of theoretically-related topics. As a result, the available data were analyzed from both perspectives, and the results of each approach are discussed here.

When conducted as a unitary study, the overall results failed to find support for the initial set of proposed hypotheses, specifically that performances on specific symptom validity tests (Rey 15 Item, the Test of Memory Malingering, the Word Memory Test, and the Forced Choice subtest from the California Verbal Learning Test, Second Edition) would demonstrate fewer scores below the established cutoff points suggesting malingering through time. Re-analysis of the symptom validity tests administered over time revealed a nearly significant difference between omission errors occurring in 2001 as compared to 2000. These results are consistent with those obtained by Grieffenstein &

Baker (2000), who found a similar trend over a 10-year span of time. However, while a statistical difference was suggested by the current results, the average numbers of omission errors for the years in question were 3 and 2, respectively, well within the normal range in both cases. It seems unlikely that a subject, already performing within the normal range on this measure, would feel the need to alter their performance further to avoid suspicion of malingering. Therefore, it is unlikely that the observed trend provides irrefutable evidence that the subjects in this study had been educated about the 15 Item Test and were making a concerted effort to perform within normal limits.

With respect to the Word Memory Test, a trend toward worse performance was found over time, with subjects' scores in 2002 falling significantly below the scores obtained in 2001 for all three trials of this test. This finding is in direct contrast to the proposed hypothesis that such scores would improve over time, had these subjects had knowledge of the intent of this test prior to assessment. The reasons for a decline in such scores are unknown, and may be related to factors that were not analyzed in this study.

It was believed that such an effect, had it been supported by the data, may have been related to subjects' growing sophistication with respect to the knowledge and use of specific symptom validity tests, or increasing familiarity with the forced-choice formats in which such tests are typically administered. The lack of support for the proposed hypotheses suggests that the patients in this sample were not influenced or assisted by such information. While the majority of subjects studied here may not have been aware that information pertaining to symptom validity testing was available to them in the popular and electronic media, the findings also suggest that such information may not have been provided to them by their attorneys or, if provided, that these subjects were unable or unwilling, for whatever reason, to make use of that information during neuropsychological assessment. While the occurrence of attorney coaching has been documented, most instances appear to be reported on a case-by-case basis (Youngjohn, 1995). Survey research of this topic (Essig et al., 2001) has suggested that approximately 75 percent of attorneys spend some time "preparing" their clients for neuropsychological assessment, with the most frequent preparation topic being the content of neuropsychological tests. The same survey indicated that approximately eight percent of attorneys provide their clients with specific test information, including how to respond to neuropsychological tests. Therefore, while these data provide some suggestion that the problem of attorney coaching is pervasive, it appears that a relatively small number of attorneys actually provide their clients with information that is specific enough to appreciably affect performance on individual symptom validity tests. However, given the limitations of survey research, the true nature of this problem remains unknown. Clearly, this is an area deserving of further study, with larger samples that span a wider range of demographic variables and include data from longer periods of time.

Additional hypotheses of this study were concerned with subject performance on specific subtests of the most recent versions of cognitive (i.e. memory) tests as a function of subjective ratings of malingering probability. That is, it was hypothesized that those subjects who were rated as malingering, based on the criteria suggested by Slick et al. (1999, Appendix A) would perform significantly below subjects who were rated as not malingering on these subtests, based on the same malingering rating criteria. Specifically, the California Verbal Learning Test, Second Edition (2000) includes a forced choice recognition subtest that is designed to function as a symptom validity measure. Similarly, the Wechsler Memory Scale, Third Edition (1997) includes subtests of recognition memory for both verbal and visual materials that are, essentially, forced choice procedures in that the respondent must respond either “yes” or “no” on tasks requiring discrimination of target items from foils. Although the overall test of significance supported the stated hypotheses, the difference between those subjects rated as Absolute malingerers and those rated as No (not malingering) was the only significant finding in all pairwise analyses of the data from both the WMS-II and the CVLT-2. The same hypothesis was proposed for subject performance on a recognition test for visual materials, the Visual Reproduction Recognition subtest of the WMS-III, which is also presented in a forced choice format. This hypothesis was also supported, with follow up tests again demonstrating a significant difference between subjects receiving malingering ratings of Absolute and No. Based on these results, it appears that the information gleaned from these subtests is useful in a comprehensive approach to assessment that includes an estimation of malingering probability. Specifically, abnormal scores on these tests appear to provide evidence of negative response bias, which is required by the criteria for malingering rating proposed by Slick et al. (1999).

Additional significant differences were found in memory test performances among the four categories of malingering probability, when these data are analyzed separately and a more lenient alpha level applied. Specifically, significant differences were found between subjects rated as Absolute and Possible malingerers on the WMS-III ARD, as well as between subjects rated as Absolute and Possible, and Probable and No on the CVLT-2 Forced Choice subtest. Therefore, not only do these subtests provide information that may distinguish absolute malingering from valid performance, they may also be helpful in situations where the subject's presentation suggests dissimulation, but is not overt. Such situations are likely to be more common, as overt malingering is a relatively infrequent event.

The results described above appear, on the surface, to provide strong evidence that the tests in question add substantively to the decision-making process in malingering detection. However, reference to the criteria used to make such determinations (Slick, 1999) indicates that the second step in the decision-making process involves the detection of negative response bias on the part of the patient, the presence of which is required for a determination of absolute malingering to be made. Since the test performances that contribute heavily to such judgments, in this case the subjects' abnormal scores on the verbal and visual memory subtests, have already been judged to be symptomatic of absolute malingering, it would follow that a correlation between these test scores and the absolute malingering ratings would be expected. In essence, this appears to be a problem of circular reasoning, in which the malingering criteria and the test performances used to make such judgments are significantly related because they are one and the same. This does not suggest that such methods are completely unsound; however, examination of the malingering rating criteria reveals that evidence of negative response bias is weighted heavily in the decision-tree approach advocated by these criteria. The counterpoint to this caveat is that negative response bias, which implies a level of performance below chance, is relatively rare and, therefore, would only be expected in those patients making a concerted effort to perform abnormally. Therefore, the finding of negative response bias may deserve its position in the decision-tree approach used in the Slick (2000) criteria, as evidence of such a bias would be expected to occur almost exclusively in patients who are consciously attempting to appear impaired.

Additional hypotheses proposed that subject performance on specific measures of psychological and emotional functioning (selected validity and Clinical Scales of the MMPI-2, the BDI-III and the BAI) would differ significantly as a function of malingering rating. Prior research (Lees-Haley, 1997; Youngjohn, Burrows & Erdal, 1995; Larrabee, 1998; Suhr et al., 1977) has demonstrated that litigating patients demonstrate MMPI-2 profiles with significant elevations on scales 1 (Hs) and 3 (Hy) at significantly greater rates than their non-litigating counterparts. The current hypotheses were supported, and are also consistent with the findings from the Lees-Haley study, in that subjects rated as absolute by the malingering criteria scored significantly higher on the Hs and Hy scales of the MMPI-2 than those subjects rated as not malingering. Significant differences were not found on the other MMPI-2 scales of interest (F, D and Pt), suggesting that malingering rating was not related to subjects' tendency to produce invalid MMPI-2 profiles or to overendorse symptoms of depression or anxiety as measured by this scale. However, Absolute malingerers did differ significantly from non-malingering subjects on the BAI. The differences in the complexity of test questions may have contributed to this difference, as the items on the BAI are simplified and ask the patient to rate their subjective experience of specific anxiety symptoms, compared to the longer and more complex MMPI-2 test items. Analysis of the BDI demonstrated a trend toward significance; but did not reach the required significance level.

A final hypothesis proposed that a significant difference would exist between malingering and non-malingering patients with respect to self-report of memory problems, as measured by the Neuropsychological Symptoms Checklist. This hypothesis was not supported, suggesting that the rate of memory complaints was similar for all subjects, regardless of referral question. Overall, the majority of subjects comprising this sample, irrespective of referral source or presenting problem, reported memory complaints.

In summary, the results of the current study failed to find strong and convincing support for the hypothesis that symptom validity test scores would demonstrate a trend toward normalization through time, ostensibly as a function of increased subject knowledge and/or preparation by plaintiff attorneys, although some trends supportive of the proposed hypotheses were seen. It was expected that such trends would emerge in the data as a direct result of the increased availability of information pertaining to symptom validity tests both in the popular media and through coaching by attorneys. While these phenomena have been documented, it appears that coaching and outright cheating on such tests are relatively low base rate events that would only emerge in studies using very large samples. These findings may be related to the relatively low numbers of subjects rated as Absolute according to the malingering rating criteria, suggesting that a relatively small portion of this sample consciously attempted to appear significantly cognitively impaired, and of that small portion, the number of subjects with prior knowledge of the specific tests could not be determined. However, it was apparent that at least one subject arrived for evaluation with some degree of preparation as, upon completing the TOMM, this individual noted that “well, at least I’m not malingering”. Unfortunately, any other subjects who may have been similarly prepared were less forthcoming about their knowledge.

Despite the lack of empirical support for the majority of the proposed hypotheses, the results of this research are of value clinically. The results underscore the relative infrequency of outright malingering, as only a small portion of the overall sample was considered to be absolutely feigning cognitive impairment. Such findings should raise caution for clinicians, as it cannot be assumed that involvement in litigation automatically predisposes a subject to malinger. In addition, the lack of strong support for any particular hypothesis in this study also cautions clinicians to avoid assigning excessive importance to any one test or assessment method, and to adopt instead the multifaceted, information gathering approach that is advocated by the Slick (2000) criteria for the determination of malingering. In summary, there does not appear to be one test instrument that will provide irrefutable evidence for the presence or absence of dissimulation or outright malingering.

The need for continued research in the area of symptom validity testing is apparent. This study did not evaluate the relationship between test performance (symptom validity or memory tests) and various demographic variables, such as education, previous litigation, number of previous neuropsychological evaluations, or gender. In addition, adequate attention was not given to those subjects rated as Probable or Possible malingerers, or those subjects whose propensity to exaggerate may be driven by internal (i.e. psychological) rather than external (monetary) sources. Finally, the nonparametric nature of the data interfered with the ability to carry out additional analyses that may have provided more comprehensive information without the burden of a very restrictive experimentwise alpha level as was required in this case. In addition, small sample sizes in some cases likely reduced the power of the analyses used. Despite these limitations, the current findings will, hopefully, provide some impetus for future research on these, and other, questions regarding the evaluation in this important area of neuropsychological practice.

4: List of References

1. Alfano, D.P., Paniak, C.E. & Finlayson, M.A. (1993). The MMPI and closed head injury: A neurocorrective approach. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, *6*(2), 111-116.

2. Alfano, D.P., Neilson, P.A. & Fink, M.P. (1993). Long-term psychosocial adjustment following head or spinal cord injury. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, *6*, 117-125.

3. Allen, L, M., III, Conder, R.L., Jr., Green, P. & Cox, D.R. (1997). *CARB '97: The Computerized Assessment of Response Bias.* Durham: CogniSyst, Inc.

4. American Psychological Association (2003, June 1). Ethical Principles for Psychologists and Code of Conduct (Standard 9.08(b)). Retrieved August 3, 2003, from http://www.apa.org/ethics/code2002.html#9_08

5. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition. Washington, DC, American Psychiatric Association, 1994.

6. Andrews, F.C. (1954). Asymptotic behavior of some rank tests for analysis of variance. As cited in S. Siegel (1956), *Nonparametric Statistics for the Behavioral Sciences,* New York: McGraw-Hill.

7. Auerbach, A.H., Scheffien, N.A. & Scholz, C.K. (1961). A questionnaire survey of the posttraumatic syndrome. *Diseases of the Nervous System*, *28*, 110-112.

8. Baker, G.A., Hanley, J.R., Jackson, H.F. (1993). Detecting the faking of amnesia: Performance differences between simulators and patients with memory impairment. *Journal of Clinical and Experimental Neuropsychology*, *15*, 668-684.

9. Barry, G. (2002). Study finds a Web of deceit on mental tests. *The Boston Globe,* June 15, 2002, p. A1.

10. Beck, A.T. (1990). *Beck Anxiety Inventory Manual,* San Antonio, TX: The Psychological Corporation.

11. Beck, A.T. (1996). *Beck Depression Inventory – Second Edition Manual,* San Antonio, TX: The Psychological Corporation.

12. Bernard, L.C. (1991). The detection of faked deficits on the Rey Auditory Verbal Learning Test: The effect of serial position. *Archives of Clinical Neuropsychology*, *6*, 81-88.

13 Bernard, L.C. (1990). Prospects for faking believable memory deficits on neuropsychological tests and the use of incentives in simulation research. *Journal of Clinical and Experimental Neuropsychology, 12*(5), 715-728.

14. Berry, D.T.R., Wetter, M.W., Baer, R.A., Youngjohn, J.R., Gass, C.G., Lamb, D.G., Franzen, M.D., MacInnes, W.D. & Bucholz, D. (1995). Over-reporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment*, *7*, 517 523.

15. Bigler, E.D., Rosa, L. & Schultz, F. (1989). Rey Auditory Verbal Learning and Rey-Osterreith Complex Figure Design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology, 45,* 277-280.

16. Binder, L.M., Rohling, M.L. & Larrabee, G.J. (1997). A review of mild head trauma. Part I: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*, *19*, 421-431.

17. Binder, L.M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology, 15*(2), 170-182.

18. Binder, L.M. (1986). Persisting symptoms after mild head injury: A review of the postconcussive syndrome. *Journal of Clinical and Experimental Neuropsychology, 4,* 25-36.

19. Bollich, A., McClain, M., Doss, C. & Black, F.W. (2002). Inconsistent memory test performance and malingering. Poster presented at the Thirteenth Annual Meeting of the International Neuropsychological Society, February 13-16, Toronto, Canada.

20. Bohnen, N., Jolles, J. & Twinjstra, A. (1992). Neuropsychological deficits in patients with persistent symptoms six months after mild head injury. *Neurosurgery, 30,* 692-696.

21. Brandt, J. (1988). Malingered amnesia. In R. Rogers (Ed.) *Clinical assessment of malingering and deception,* pp. 65-83. New York: Guilford Press.

22. Brooks, N., Campsie, L, Symington, C., Beattie, A. & McKinlay, W. (1986). The five year outcome of sever blunt-head injury. *Acta Neurochirchica*, *34*, 57-70.

23. Brulot, M.M., Strauss, E. & Spellacy, F. (1997). Validity of the Minnesota Multiphasic Personality Inventory-2 correction factors for use with patients with suspected head injury. *The Clinical Neuropsychologist, 11*(4), 391-401.

24. Butcher, J.N & Williams, C.L. (1992). *Essentials of MMPI-2 and MMPI-A Interpretation.* Minneapolis: University of Minnesota Press.

25. Butcher, J. N., Dahlstrom, W.G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory-2 (MMPI-2): Manual for administration and scoring.* Minneapolis: University of Minnesota Press.

26. Caine, E.D. (1986). The neuropsychology of depression: The pseudodementia syndrome. In Grant, I. & Adams, K.A. (Eds.) *Neuropsychological Assessment of Neuropsychiatric Disorders.* Oxford: Oxford University Press.

27. Cercy, S.P., Schretlen, D.J. & Brandt, J. (1997). Simulated amnesia and the pseudo-memory phenomenon. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception – Second Edition,* pp. 85-107. New York: Guilford Publications, Inc.

28. Coleman, R.D, Rapport, L.J., Millis, S.R., Richer, J.H. & Farchione, T.J. (1998). Effects of coaching on detection of malingering on the California Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*, *20*(2), 201-210.

29. Connor, D.J, Drake, A.I., Bondi, M.W. & Delis, D.C. (1997). *Detection of Feigned Cognitive Impairments in Patients with a History of Mild to Severe Closed Head Injury.* Presented at the meetings of the American Academy of Neurology, Boston, MA. Abstract obtained from *The California Verbal Learning Test – Second Edition, Adult Version,* San Antonio: The Psychological Corporation.

30. Cronholm, B. & Ottonsson, J. (1961). Memory functions in endogenous depression. *Archives of General Psychiatry*, 5, 193-197.

31. Crosson, B., Novack, T.A., Trennery, M.R. & Craig, P.L. (1988). California Verbal Learning Test (CVLT) performance in severely head-injured and neurologically normal adult males. *Journal of Clinical and Experimental Neuropsychology, 10*(6), 754-768.

32. Delis, D.C., Kramer, J.H., Kaplan, E. & Ober, B.A. (1987). *California Verbal Learning Test – First Edition, Adult Version.* San Antonio: The Psychological Corporation.

33. Delis, D.C., Kramer, J.H., Kaplan, E. & Ober, B.A. (2000). *California Verbal Learning Test – Second Edition, Adult Version.* San Antonio: The Psychological Corporation.

34. Dikmen, S.S., Machamer, J.E., Winn, H.R. & Temkin, N. R. (1995). Neuropsychological outcome at 1-year post head injury. *Neuropsychology*, *9*(1), 80-90.

35. Dikmen, S.S., Machamer, J., Temkin, N. & McLean, A. (1990). Neuropsychological recovery in patients with moderate to severe head injury: 2 year follow-up. *Journal of Clinical and Experimental Neuropsychology, 12*(4), 507-519.

36. Dikmen, S.S., Temkin, N., McLean, A., Wyler, A. & Machamer, J. (1987). Memory and head injury severity. *Journal of Neurology, Neurosurgery and Psychiatry, 50,* 1613-1618.

37. Dikmen, S.S., McLean, A., Jr., Temkin, N. & Wyler, A. (1986a). Neuropsychologic outcome at one-month post injury. *Archives of Physical Medicine and Rehabilitation, 67,* 507-513.

38. Dikmen, S.S., McLean, A. & Temkin, N. (1986b). Neuropsychological and psychosocial consequences of minor head injury. *Journal of Neurology, Neurosurgery and Psychiatry, 49,* 1227-1232.

39. Dunn, J.T. & Lees-Haley, P.R. (1995). The MMPI-2 correction factor for closed head injury: A caveat for forensic cases. *Assessment, 2,* 47-51.

40. Dywan, J., Segalowitz, S.J. Henderson, K. & Jacoby, L. (1993). Memory for source after traumatic brain injury. *Brain and Cognition*, *21*, 20-43.

41. Essig, S.M., Mittenberg, W., Petersen, R.S., Strauman, S. & Cooper, J.T. (2001). Practices in forensic Neuropsychology: Perspectives of neuropsychologists and trial attorneys. *Archives of Clinical Neuropsychology, 16,* 271-291.

42. Fann, J.R., Katon, W.J., Uomoto, J.M. & Esselman, P.C. (1995). Psychiatric disorders and functional disability in outpatients with traumatic brain injuries. *American Journal of Psychiatry*, *152*, 1493-1499.

43. Fee, C.R. & Rutherford, W.H. (1988). A study of the effect of legal settlement on post-concussion syndrome. *Archives of Emergency Medicine, 5,* 12-17.

44. Filley, C.M. (1995). *Neurobehavioral Anatomy*. Niwot: University of Colorado Press.

45. Frederick, R.I. & Foster, H.G., Jr. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3,* 596-602.

46. Frederick, R.I., Sarfaty, S.D., Johnston, J.D. & Powell, J (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology*, *8*, 118-125.

47. Gass, C. S. (1996). MMPI-2 variables in attention and memory performance. *Psychological Assessment*, *8*, 135-138.

48. Gouvier, W.D., Uddo-Crane, M. & Brown, L.M. (1988). Base rates of post concussional symptoms. *Archives of Clinical Neuropsychology, 3,* 273-278.

49. Gouvier, W.D., Cubic, B., Jones, G., Brantley, P. & Cutlip, Q. (1992). Postconcussion symptoms and daily stressing normal and head injured college populations. *Archives of Clinical Neuropsychology, 7,* 192-211.

50. Green, P., Allen, L.M. & Astner, K. (1996). *The Word Memory Test: A user's guide to the oral and computer-administered forms*, US Version 1.1 Durham: CogniSyst, Inc.

51. Green, P., Iverson, G.L. & Allen, L. (1999). Detecting malingering in head injury litigation with the Word Memory Test. *Brain Injury, 13,* 813-819.

52. Grieffenstein, M.F. & Baker, W.J. (2002). The Test of Time: Does the Rey 15 Item Test still work? Poster presented at the 22nd annual conference of the National Academy of Neuropsychology, October 9-13, Miami, FL.

53. Grieffenstein, M.F., Baker, W.J. & Gola, T. (1994). Validation of malingered amnesia with a large clinical sample. *Psychological Assessment*, *6*, 218-224.

54. Greiffenstein, M.F., Gola, T. & Baker, W.J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist*, *9*, 230-240.

55. Haines, M.E. & Norris, M.P. (2001). Comparing student and patient simulated malingerers' performance on standard neuropsychological measures to detect feigned cognitive deficits., *The Clinical Neuropsychologist, 15*(2), 171-182.

56. Hathaway, S.R. & McKinley, J.C. (1943). *The Minnesota Multiphasic Personality Inventory Manual*, New York: Psychological Corporation.

57. Hathaway, S.R. & McKinley, J.C. (1989). *The Minnesota Multiphasic Personality Inventory – Second Edition Manual*, The University of Minnesota Press.

58. Heaton, R.K., Smith, H.H., Jr., Lehman, R.A. & Vogt, A.J. (1978). Prospects for faking believable deficits on neuropsychological testing. *Journal of Consulting and Clinical Psychology*, *46*, 892-900.

59. Hiscock, M & Hiscock, C.K. (1989). Refining the forced choice method for the detection of malingering. *Journal of Clinical and Experimental Neuropsychology*, *11*, 967-974.

60. Iverson, G.L., Franzen, M.D. & McCracken, L.M. (1991). Evaluation of an objective assessment technique for the detection of malingering memory deficits. *Law and Human Behavior, 16,* 667-676.

61. Iverson, G.L. & Binder, L.M. (2000). Detecting exaggeration and malingering in neuropsychological assessment. *Journal of Head Trauma Rehabilitation*, *15*(2), 829-858.

62. Iverson, G.L., King, R.J., Scott, J.G. & Adams, R.L. (2001). Cognitive complaints in litigating patients with head injuries or chronic pain. *Journal of Forensic Neuropsychology*, *2*(1), 19-30.

63. Iverson, G.L. & McCracken, L.M. (1997). “Postconcussive” symptoms in persons with chronic pain. *Brain Injury*, *11*, 783-790.

64. Iverson, G.L. & Franzen, M.D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology*, *18*(1), 38-51.

65. Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513-541.

66. Johnson, J.L. & Lesniak-Karpiak, K. (1997). The effects of warning on malingering on memory and motor tasks in college samples. *Archives of Clinical Neuropsychology, 12*, 231-238.

67. Kareken, D.A., Moberg, P.J. & Gur, R.C. (1996). Proactive inhibition and semantic organization: Relationship with verbal memory in patients with schizophrenia. *Journal of the International Neuropsychological Society*, *2*, 486-493.

68. Killgore, W.D.S. & DellaPietra, L. (2000a). Item response bias on the Logical Memory delayed recognition subtest of the Wechsler Memory Scale-III. *Psychological Reports, 86* (3 Pt. 1), 851-857.

69. Killgore, W.D.S. & DellaPietra, L. (2000b). Using the WMS-III to detect malingering: Empirical validation of the Rarely Missed Index (RMI). *Journal of Clinical and Experimental Neuropsychology, 22*(6), 761-771.

70. King, D.H. & Caine, E.D. (1996). Cognitive impairment and major depression: Beyond the pseudodementia syndrome. In: I. Grant & K. Adams (eds.), *Neuropsychological Assessment of Neuropsychiatric Disorders* (2nd Edition). New York: Oxford University Press.

71. Kolpan, K.I. (1996). Your life is an open book. *Journal of Head Trauma Rehabilitation*, *11*(2), 102-103.

72. Larrabee, G.J. (2000). Preface. *Journal of Head Trauma Rehabilitation*, 15(2), v viii.

73. Larrabee, G.J. (1998). Somatic malingering on the MMPI and MMPI-2 in litigating subjects. *The Clinical Neuropsychologist*, *12,* 179-188.

74. Larrabee, G.J. (1990). Cautions in the use of neuropsychological evaluation in legal settings. *Neuropsychology*, *4*, 239-247.

75. Lee, G.P., Loring, D.W. & Martin, R.C. (1992). Rey's 15-item visual memory test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment*, *4*, 43-46.

75. Lees-Haley, P.R., Fox, D.D. & Courtney, J.C. (2001). A comparison of complaints by mild brain injury claimants and other claimants describing subjective experiences immediately following their injury. *Archives of Clinical Neuropsychology, 16,* 689-695.

76. Lees-Haley, P.R. & Brown, R.S. (1993). Neuropsychological complaint base rates of 170 personal injury claimants. *Archives of Clinical Neuropsychology*, *8*, 203 209.

77. Lees-Haley (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology, 53*(7), 745-755.

78. Lees-Haley, P.R., English, L.T. & Glenn, W.J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *The Clinical Neuropsychologist, 68,* 203 210.

79. Levin, H., Mattis, S., Ruff, R., Eisenberg, H., Marshall, L., Tabador, K., High, W. & Frankowski, R. (1987). Neurobehavioral outcome following minor head injury: A three-center study. *Journal of Neurosurgery, 66,* 234-243.

80. Lezak, M.D. (1995). *Neuropsychological Assessment: Third Edition.* Oxford: Oxford University Press.

81. Luria, A.R. (1981). *Higher cortical functions in man.* New York: Basic Books.

82. Maddocks, D. & Saling, M. (1996). Neuropsychological deficits following concussion. *Brain Injury, 10,* 99-104.

83. Massman, P.J., Delis, D, C., Butters, N., Dupont, R.M. & Griffin, J.C. (1992). The subcortical dysfunction hypothesis of memory deficits in depression: Neuropsychological validation in a subgroup of patients. *Journal of Clinical and Experimental Neuropsychology, 14,* 687-706.

84. McClain, M. & Black, F.W. (2002). The effect of mood on symptom validity test performance. Poster presented at the Thirteenth Annual Meeting of the International Neuropsychological Society, February 13-16, Toronto, Canada.

85. Miller, H. (1961). Accident neurosis. *British Medical Journal*, i, 919.

86. Millis, S.R., Putnam, S.H., Adams, K.M. & Ricker, J.H. (1995). The California Verbal Learning Test in the detection of incomplete effort in neuropsychological evaluation. *Psychological Assessment*, *7*, 463-471.

87. Mills, S.R. & Kler, S. (1995). Limitations of the Rey Fifteen-Item Test in the detection of malingering. *The Clinical Neuropsychologist*, *9*(3), 241-244.

88. Mittenberg, W. & Strauman, S. (2000). Diagnosis of mild head injury and the postconcussion syndrome. *Journal of Head Trauma Rehabilitation*, *15*(2), 783-791.

89. Mittenberg, W., Tremont, G. & Zielinski, R.E. (1996). Cognitive-behavioral presentation of postconcussion syndrome. *Archives of Clinical Neuropsychology, 11,* 139-145.

90. Mittenberg, W., Azrin, R., Millsaps, C. & Heilbronner, R. L. (1993). Identification of malingered head injury on the Wechsler Memory Scale-Revised. *Psychological Assessment, 5,* 34-40.

91. Morley, G.K., Haxby, J.V. & Lundgren, S.L. (1980). Memory, aging and dementia. In G.J. Maletta & F.J. Pirozzolo (Eds.), *The aging nervous system*. New York: Praeger.

92. Mutter, S.A., Howard, D.V. & Wiggs, C.L. (1990). Performance on direct and indirect tests of memory after mild closed head injury. *Cognitive Neuropsychology*, *7*, 329-346.

93. Nies, K.J. & Sweet, J.L. (1994). Neuropsychological assessment and malingering: A critical review of past and present strategies. *Archives of Clinical Neuropsychology, 9,* 501-552.

94. Nott, P.N. & Fleminger, J.J. (1975). Presenile dementia: The difficulties of early diagnosis. *Acta Psychologica Scandinavica*, *51*, 210-217.

95. Novack, T.A., Kofoed, B.A. & Crosson, B. (1995). Sequential performance on the California Verbal Learning Test following traumatic brain injury. *The Clinical Neuropsychologist, 9,* 38-43.

96. O'Connor, M.G. & Morin, M.D. (1998). Amnesic syndromes. In Snyder, P.J. & Nussbaum, P.D. (Eds.) *Clinical Neuropsychology: A pocket handbook for assessment.* Washington, DC: American Psychological Association, p. 426-449.

97. Pankratz, L., Fausti, S. & Peed, S. (1975). A forced-choice technique to evaluate deafness in the hysterical or malingering patient. *Journal of Consulting and Clinical Psychology*, 43, 421-422.

98. Parkin, A.J. & Russo, R. (1990). Implicit and explicit memory and the automatic /effortful distinction. *European Journal of Cognitive Psychology*, *2*, 71-80.

99. Psychological Assessment Resources (1983). *Neuropsychological Symptoms Checklist.* Odessa, FL: PAR.

100. Rees, L.M., Tombaugh, T.N. & Boulay, L. (2001). Depression and the Test of Memory Malingering. *Archives of Clinical Neuropsychology*, *16*(5), 501-506.

101. Reitan, R.M. & Wolfson, D. (1997). Emotional disturbances and their interactions with neuropsychological deficits. *Neuropsychology Review*, *7*, 3-19.

102. Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaries de France.

103. Rogers, R., Sewell, K.W. & Goldstein, A. (1994). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior, 18,* 543-552.

104. Rogers, R., Harrell, E.H., & Liff, C.D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review, 13,* 255-274.

105. Rogers, R. (1997). Researching dissimulation. In R. Rogers (Ed.) *Clinical Assessment of Malingering and Deception – Second Edition,* pp. 398-426. New York: Guilford Publications, Inc.

106. Rose, S.F., Hall, S. & Szalda-Petree, A.D. (1995). Computerized Portland Digit Recognition Test: The measurement of response latency improves the detection of malingering. *The Clinical Neuropsychologist*, *9*, 124-134.

107. Ruiz, M.A., Drake, E.B., Glass, A., Marcotte, D. & van Gorp, W.G. (2002). Trying to beat the system: Misuse of the Internet to assist in avoiding the detection of psychological symptom dissimulation. *Professional Psychology: Research and Practice, 33*(3), 294-299.

108. Sanders, R.E., Murphy, M.D., Schmitt, F.A. & Walsh, K.K. (1980). Age differences in free recall rehearsal strategies. *Journal of Gerontology*, *35*, 550-558.

109. Schacter, D.L. (1986). Amnesia and Crime: How much do we really know? *The American Psychologist, 41*(3), 286-295.

110. Schretlen, D., Brandt, J. & Krafft, L. (1991). Some caveats in using the Rey 15 Item Memory Test to detect malingered amnesia. *Psychological Assessment*, *3*(4), 667-672.

111. Schmitter-Edgecomb, M., Marks, W. & Fahy, J.F. (1993). Semantic priming following severe closed head trauma: Automatic and attentional processes. *Neuropsychology*, 7, 136-148.

112. Schmitter-Edgecomb, M. (1996). Effects of traumatic brain injury on cognitive performance :An attentional resource hypothesis in search of data. *Journal of Head Trauma Rehabilitation*, *11*(2), 17-30.

113. Selnes, O.A., Jacobson, L. & Machado, A.M (1991). Normative data for a brief neuropsychological screening battery. *Perceptual and Motor Skills, 73,* 539-550.

114. Sherer, M., Bergloff, P., Levin, E., High, W.M. Jr., Oden, K.E. & Nick, T.G. (1998a). Impaired awareness and employment outcome after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *13*(5), 52-61.

115. Sherer, M., Boake, C., Levin, E., Silver, B.V., Ringholz, G. & High, W.M. Jr. (1998b). Characteristics of impaired awareness after traumatic brain injury. *Journal of the International Neuropsychological Society*, *4*, 380-387.

116. Sherer, M., Madison, C.F. & Hannay, H.J. (2000). A review of outcome after moderate and severe closed head injury with an introduction to life care planning. *Journal of Head Trauma Rehabilitation*, *15*(2), 767-782.

117. Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences,* New York: McGraw Hill.

118. Slick, D.J., Iverson, G.L. & Green P. (2000). California Verbal Learning Test indicators of suboptimal performance in a sample of head-injury litigants. *Journal of Clinical and Experimental Neuropsychology*, *22*(5), 569-579.

119. Slick, D.J., Sherman, E.M.S. & Iverson, G.L. (1999). Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *The Clinical Neuropsychologist, 13*(4), 545-561.

120. Slick, D., Hopp, G., Strauss, E. & Spellacy, F. (1996). Victoria Symptom Validity Test: Efficiency for detecting feigned memory impairment and relationship to neuropsychological tests and MMPI-2 validity scales. *Journal of Clinical and Experimental Neuropsychology, 18,* 911-922.

121. Slick, D., Hopp, G., Strauss, E., Hunter, M. & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's Forced-Choice Memory Test. *Journal of Clinical and Experimental Neuropsychology*, *16*(3), 472 481.

122. Smith, J.B., Kiloh, G.L., Ratnavale, G.S. (1976). The investigation of dementia: The results of 100 consecutive admissions. *Medical Journal of Australia*, *2*, 403 405.

123. Squire, L. (1987). *Memory and Brain.* New York: Oxford.

124. Suhr, J, Tranel, D., Wefel, J. & Barrash, J. (1997). Memory performance after head injury: Contributions of malingering, litigation status, psychological factors, and medication use. *Journal of Clinical and Experimental Neuropsychology*, *19*(4), 500-514.

125. Suhr. J. A. & Gunstad, J. (2000). The effects of coaching on the sensitivity and specificity of malingering measures. *Archives of Clinical Neuropsychology, 15*(5), 415-424.

126. Sullivan, K., Keane, B. & Deffenti, C. (2001). Malingering on the RAVLT Part I.: Deterrence strategies. *Archives of Clinical Neuropsychology, 16,* 627-641.

127. Teasdale, G. & Jennett, B. (1974). Assessment of coma and impaired consciousness. *Lancet, ii,* 81-84.

128. The Psychological Corporation: *WAIS-III/WMS-III Technical Manual*. San Antonio: The Psychological Corporation, 1997.

129. Tombaugh, T.N. (1996). *The Test of Memory Malingering*. North Tonawanda: Multi-Health Systems, Inc.

130. Trahan, D.E., Ross, C.E. & Trahan, S.L. (2001). Relationships among postconcussional-type symptoms, depression, and anxiety in neurologically normal young adults and victims of mild brain injury. *Archives of Clinical Neuropsychology, 16*(5), 435-445.

131. Trueblood, W. & Schmidt, M. (1993). Malingering and other validity considerations in the neuropsychological evaluation of mild head injury. *Journal of Clinical and Experimental Neuropsychology*, *15*, 578-590.

132. Wechsler, D. (1997). *Wechsler Memory Scale – Third Edition Administration and Scoring Manual.* San Antonio, TX: The Psychological Corporation.

133. Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87-95.

134. Weingartner, H., Gold, P., Ballenger, J.D. (1981). Effects of vasopressin on human memory functions. *Science*, *211*, 601-603.

135. Williams, J.M. (1998). The malingering of memory disorder. In Reynolds, C.R. (Ed.) *Detection of Malingering During Head Injury Litigation.* New York: Plenum Press.

136. Wong, J.L., Regennitter, R.P. & Barrios, F. (1994). Base rate and simulated symptoms of mild head injury among normals. *Archives of Clinical Neuropsychology*, *9*, 411-425.

137. World Health Organization. *International Statistical Classification of Diseases and Related Health Problems.* 10th edition. Geneva, Switzerland: World Health Organization, 1992.

138. Youngjohn, J.R., Lees-Haley, P.R. & Binder, L.M. (1999). Comment: Warning malingerers produces more sophisticated malingering. *Archives of Clinical Neuropsychology, 14*(6), 511-515.

139. Youngjohn, J.R., Burrows, L. & Erdal, K. (1995a). Brain damage or compensation neurosis? The controversial post-concussion syndrome. *The Clinical Neuropsychologist, 9,* 112-123.

140. Youngjohn, J.R. (1995b). Confirmed attorney coaching prior to neuropsychological evaluation. *Assessment, 2,* 279-283.

141. Zakzanis, K.K., Leach, L. & Kaplan, E. (1998). On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry, Neuropsychology and Behavioral Neurology*, *11*, 111-119.

Appendix A: Clinical Assessment of Malingering Probability

A. Presence of substantial external incentive

B. Evidence from neuropsychological functioning

    1. Definite negative response bias

    2. Probable response bias

    3. Discrepancy between the test data and known patterns of brain function

    4. Discrepancy between test data and observed data

    5. Discrepancy between test data and reliable collateral reports

    6. Discrepancy between test data and documented background history

C. Evidence from patient self-report

    1. Self-report history is discrepant with documented history

    2. Self-reported symptoms are discrepant with known patterns of brain functioning

    3. Self-reported symptoms are discrepant with behavioral observations

    4. Self-reported symptoms are discrepant with information obtained from collateral informants.

    5. Evidence of exaggerated or fabricated psychological (neurocognitive) dysfunction.

D. Behaviors meeting criteria from groups B and C are not fully accounted for by known Psychiatric, Neurological or Developmental factors.

Absolute Malingering: Meets criterion A and Criterion B1 and Criterion D.

Probable Malingering: Meets Criterion A and two or more B Criteria (excluding B1) and one or more C Criteria.

Possible Malingering: Meets Criterion A <u>and</u> one or more C Criteria, but does not meet Criterion D; or, meets all criteria for Definite or Probable Malingering, but does not meet Criterion D.

Not Malingering: May meet Criterion A, but does not meet any B, C, or D Criteria.

Vita

Maryellen McClain, M.S., M.A

Date of birth: January 16, 1965

Place of birth: Sellersville, Pennsylvania

Education:

Albright College, Reading, Pennsylvania – Bachelor of Science, Psychobiology, May 1986

Drexel University, Philadelphia, Pennsylvania – Master of Science, Neuropsychology, December 1992

MCP Hahnemann University, Philadelphia, Pennsylvania – Master of Arts, Clinical Psychology, December, 1998

Tulane University Health Sciences Center, New Orleans, Louisiana – Clinical Psychology Internship, with specialization in Neuropsychology, July 1, 2000 to July 31, 2001

Tulane University Health Sciences Center, New Orleans, Louisiana – Postdoctoral Fellowship in Clinical Neuropsychology, August 1, 2001 to July 31, 2003

Teaching and Professional Experience:

Psychometrist – Magee Rehabilitation Hospital (1989-1996)

Teaching Assistant – Neuropsychological Assessment, Hahnemann University (1994-1996)

Teaching Assistant – Psychological Assessment, MCP Hahnemann University (1997-1998)